Q. J. R. Meteorol. Soc. (1998), **124**, pp. 1669–1694

Efficient methods to assimilate remotely sensed data based on information content

By J. JOINER* and A. M. DA SILVA
NASA/Goddard Space Flight Center, USA

(Received 20 November 1996; revised 25 June 1997)

SUMMARY

Two basic approaches have evolved to utilize measurements of radiance (i.e. thermal or scattered solar radiation) by satellite-borne instruments in data assimilation systems: radiances (raw or cloud-corrected) may be assimilated directly, or they may be pre-processed to retrieve geophysical parameters for subsequent assimilation. The retrieval process is often ill-posed, and therefore requires the use of prior information to constrain the solution. For example, temperature and humidity profiles retrieved using radiances from nadir-viewing infrared and microwave sounders often incorporate prior information in the form of climatology or forecasts. The use of prior information presents difficulties when assimilating retrievals. Here we present methods to remove prior information from retrievals in order to achieve a more consistent assimilation of the data. In addition, these methods can be used as a data compression device, which can reduce the amount of computation required by some analysis systems compared with radiance assimilation. The methods are implemented and compared in a one-dimensional assimilation system using simulated data from current and future infrared-temperature profiling instruments.

KEYWORDS: Data assimilation Retrieval techniques Satellite data

1. INTRODUCTION

In operational data assimilation systems (DASs), satellite data and conventional meteorological data are combined with forecasts from a numerical model to produce a global, physically consistent estimate of the atmospheric state. For many years, the primary purpose of data assimilation has been to provide initial conditions for numerical weather prediction. More recently, data assimilation has been recognized as an important tool for climate and earth system studies (e.g. Bengtsson and Shukla 1988). Several centres are currently engaged in the preparation of multi-year assimilated data sets produced with a single, consistent DAS (Schubert *et al.* 1993; Gibson *et al.* 1994; Kalnay *et al.* 1995).

In this paper, we focus on methods to assimilate temperature information from nadir-viewing satellite-borne instruments that measure radiances (i.e. thermal or scattered/reflected radiation from earth) in different spectral bands. Although this represents only a single data type, the concepts that will be developed are applicable to the assimilation of other data types such as precipitation, precipitable water vapour, or wind speed information derived from remote-sensing instruments.

Two basic approaches have been used to incorporate measurements from remote-sounding instruments in DASs. The more traditional approach is to assimilate geophysical products retrieved from radiances. In this approach, radiances from an individual sounding are processed off-line to produce a set of geophysical parameters, such as 1D profiles of temperature or humidity, that are used in a DAS. The retrieval error covariance is required for retrieval assimilation and is usually statistically derived. Retrieval errors are often assumed to be horizontally isotropic, stationary, and uncorrelated with forecast errors. In many cases, some or all of these assumptions are incorrect, as discussed by Daley (1993) and as shown by Sullivan *et al.* (1993). In particular, retrieval errors are often significantly correlated with forecast errors. This correlation results in part from the fact that the retrieval is a nonlinear estimation process that is often ill-posed and therefore requires the use of prior

* Corresponding author: Data Assimilation Office, Code 910.3, NASA/GSFC, Greenbelt, MD 20771, USA. e-mail: joiner@hera.gsfc.nasa.gov.

information. The errors of the prior information incorporated into the retrieval are likely to be correlated with forecast errors. For example, in an interactive system, forecasts are used as prior information for the retrievals (e.g. Susskind and Pfaendtner 1989; Goldberg *et al.* 1993; Eyre *et al.* 1993). Regression, pattern recognition (e.g. Chédin *et al.* 1985), and neural net (e.g. Escobar-Munoz *et al.* 1993) retrievals also contain prior information in the form of an ensemble of profiles used to derive statistical relationships in the algorithms.

The second approach for utilizing remotely sensed data in a DAS involves directly assimilating radiances. This approach requires the specification of a radiance error covariance. The assumptions that observation errors are state-independent, stationary, and uncorrelated with the forecast are more justified for radiances than for retrievals (e.g. Eyre *et al.* 1993). Preprocessing of the radiances, such as removing the effects of clouds, will create non-stationary radiance errors. In this situation, a state-dependent radiance error covariance may be used (Derber and Wu 1996). Radiance assimilation is computationally feasible with current instruments and analysis schemes. However, new approaches may be necessary to make use of the full information content of future high-spectral-resolution sounding instruments in some assimilation systems.

In this paper methods are presented to remove prior information from retrievals by the partial eigen-decomposition (PED) or partial singular-value decomposition of different retrieval operators. Coefficients of eigenvectors or singular-vectors are assimilated as a new data type. Coefficients that are not well determined by the radiances may be truncated with no significant loss of information content in the DAS. This compact truncated-mode representation of the information content in radiances may result in a significant reduction in the amount of computation required by a statistical analysis system, such as the Physical-space Statistical Analysis System (PSAS) under development at the NASA/Goddard Data Assimilation Office (da Silva *et al.* 1995; Guo and da Silva 1995). Similar approaches involving singular-value decomposition (SVD) and eigen-decomposition of retrieval operators have been used for many years in 1D retrieval algorithms (e.g. Twomey 1974; Smith and Woolf 1976; Thompson 1992). However, this type of approach has not been previously used, to our knowledge, in a data assimilation context.

Following an outline of the notation used in the remainder of the paper, methods for assimilating remotely sensed data are briefly reviewed. The specific case of radiance assimilation is described in section 4. In section 5 we provide a description of physical-space retrieval assimilation. This forms a foundation for the development of phase-space retrieval assimilation in section 6. In section 7 we compare data assimilation methods in a 1D DAS using simulated data from current and advanced infrared-temperature sounding instruments. Concluding remarks and future plans appear in section 8.

2. NOTATION

Before proceeding, we briefly outline the notation that will be used. Different notations have evolved in data assimilation and retrieval literature. For example, K is used in data assimilation literature to mean the Kalman gain matrix (e.g. Daley 1991). In retrieval literature, K is commonly used to represent the Jacobian of the radiative-transfer equation (e.g. Rodgers 1976). Here we have attempted to follow the notation of Rodgers (1990) wherever possible, in combination with notations used in data assimilation literature. We have also introduced some less conventional notations. The following are symbols adopted for general variables:

$\mathbf{w}$	state variable
$\boldsymbol{\epsilon}$	error
$\mathbf{P}$	forecast-error covariance
$\mathbf{R}$	observation-error covariance
h	general observation operator mapping state variable to the observation space
H	linearized version of the observation operator
$\mathcal{I}$	interpolation operator that maps state variables on the analysis grid to observation locations

The following symbols will be used as subscripts and superscripts for the above quantities:

T	transpose of a matrix
o	observed
t	truth
f	forecast
a	analysis
p	retrieval prior information
g	grid

The following symbols represent different types of observations and may also appear as superscripts or subscripts (in italic):

$\mathbf{y}$	radiance
$\mathbf{z}$	physical-space retrieval
$\boldsymbol{\alpha}$	phase-space retrieval

For example, $\mathbf{w}^{\mathrm{t}}$ denotes the true state, $\mathbf{w}^{\mathrm{f}}$ denotes the forecast state, $\boldsymbol{\epsilon}^{\mathrm{f}} = \mathbf{w}^{\mathrm{f}} - \mathbf{w}^{\mathrm{t}}$ denotes the forecast error, and $\mathbf{P}^{\mathrm{f}} = \langle \boldsymbol{\epsilon}^{\mathrm{f}} (\boldsymbol{\epsilon}^{\mathrm{f}})^{\mathrm{T}} \rangle$ is the forecast-error covariance.

3. Statistical analysis

The objective of statistical interpolation is to produce an optimal estimate of the atmospheric state, given a set of observations and a first guess—usually in the form of a short-term forecast. In the variational framework (e.g. Lorenc 1986; Talagrand 1988), this can be accomplished by minimizing the likelihood functional

$$J(\mathbf{w}) = \left(\mathbf{w} - \mathbf{w}^{\mathrm{f}}\right)^{\mathrm{T}} (\mathbf{P}^{\mathrm{f}})^{-1} \left(\mathbf{w} - \mathbf{w}^{\mathrm{f}}\right) + \left(\mathbf{w}^{\mathrm{o}} - \mathrm{h}(\mathbf{w})\right)^{\mathrm{T}} (\mathbf{R}^{\mathrm{o}})^{-1} \left(\mathbf{w}^{\mathrm{o}} - \mathrm{h}(\mathbf{w})\right), \qquad (1)$$

where $\mathbf{w} \in \Re^{n_{\mathrm{g}}}$ is a vector of length n_{g} representing the 3D atmospheric state, $\mathbf{w}^{\mathrm{f}} \in \Re^{n_{\mathrm{g}}}$ is the forecast, $\mathbf{w}^{\mathrm{o}} \in \Re^{n_{\mathrm{o}}}$ is the observation vector of length n_{o}, and $\mathrm{h}(\mathbf{w})$ is the observation operator that maps the 3D atmospheric state to the observation space. The first term on the right-hand side (r.h.s.) of (1) is weighted by the inverse of the forecast-error covariance $\mathbf{P}^{\mathrm{f}} \in \Re^{n_{\mathrm{g}}} \times \Re^{n_{\mathrm{g}}}$, and the second term is weighted by the inverse of the observation error covariance $\mathbf{R}^{\mathrm{o}} \in \Re^{n_{\mathrm{o}}} \times \Re^{n_{\mathrm{o}}}$. Provided these covariances are specified correctly, the analysis state obtained by minimizing $J(\mathbf{w})$ is the mode of the conditional probability density function $\mathrm{p}(\mathbf{w}|\mathbf{w}^{\mathrm{f}} \bigcup \mathbf{y})$ (Jazwinski 1970) and is derived from a maximum-likelihood principle assuming that forecast and observation errors are unbiased, normally distributed, and uncorrelated with each other. Algorithms for the minimization of (1) have been described in Navon and Legler (1987) and implemented in global data assimilation systems by Courtier *et al.* (1993) and Parrish and Derber (1992). Next we outline the numerical

algorithm incorporated in the PSAS, as it is important to concepts developed in subsequent sections.

The observation operator h(**w**) is in general nonlinear. Therefore the minimum of $J(\mathbf{w})$ can be obtained by a quasi-Newton iteration of the form

$$\begin{aligned}\mathbf{w}_{i+1} &= \mathbf{w}^{\mathrm{f}} + \mathbf{P}^{\mathrm{f}}\mathbf{H}_i^{\mathrm{T}}\left(\mathbf{H}_i\mathbf{P}^{\mathrm{f}}\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}^{\mathrm{o}}\right)^{-1}\left[\mathbf{w}^{\mathrm{o}} - \mathrm{h}(\mathbf{w}_i) + \mathbf{H}_i(\mathbf{w}_i - \mathbf{w}^{\mathrm{f}})\right] \\ &= \mathbf{w}^{\mathrm{f}} + \left(\mathbf{H}_i^{\mathrm{T}}(\mathbf{R}^{\mathrm{o}})^{-1}\mathbf{H}_i + (\mathbf{P}^{\mathrm{f}})^{-1}\right)^{-1}\mathbf{H}_i^{\mathrm{T}}(\mathbf{R}^{\mathrm{o}})^{-1}\left[\mathbf{w}^{\mathrm{o}} - \mathrm{h}(\mathbf{w}_i) + \mathbf{H}_i(\mathbf{w}_i - \mathbf{w}^{\mathrm{f}})\right] ,\end{aligned} \tag{2}$$

where

$$\mathbf{H}_i = \left.\frac{\partial \mathrm{h}(\mathbf{w})}{\partial \mathbf{w}}\right|_{\mathbf{w}=\mathbf{w}_i} . \tag{3}$$

The analysis vector, $\mathbf{w}^{\mathrm{a}}$, that minimizes $J(\mathbf{w})$ is given by

$$\mathbf{w}^{\mathrm{a}} = \lim_{i\to\infty} \mathbf{w}_i . \tag{4}$$

In a PSAS, a $n_{\mathrm{o}} \times n_{\mathrm{o}}$ system of equations is first solved at observation locations for the vector $\mathbf{x}_i \in \Re^{n_{\mathrm{o}}}$ using a conjugate gradient algorithm (da Silva *et al.* 1995; Guo and da Silva 1995), i.e.

$$\left(\mathbf{H}_i\mathbf{P}^{\mathrm{f}}\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}^{\mathrm{o}}\right)\mathbf{x}_i = \mathbf{w}^{\mathrm{o}} - \mathrm{h}(\mathbf{w}_i) + \mathbf{H}_i\left(\mathbf{w}_i - \mathbf{w}^{\mathrm{f}}\right) . \tag{5}$$

The first term on the left-hand side of (5) will be referred to as the innovation covariance. The state at iteration $(i + 1)$ is updated by an additional matrix-vector multiplication, i.e.

$$\mathbf{w}_{i+1} = \mathbf{w}^{\mathrm{f}} + \mathbf{P}^{\mathrm{f}}\mathbf{H}_i^{\mathrm{T}}\mathbf{x}_i . \tag{6}$$

The computation required for the solution of the linear system (5) is approximately $\mathrm{O}(N_{\mathrm{cg}}n_{\mathrm{o}}^2)$, where N_{cg} is the number of iterations of the conjugate gradient algorithm. N_{cg} depends upon the conditioning of the innovation covariance. The matrix-vector multiplication in (6) requires $\mathrm{O}(n_{\mathrm{g}}n_{\mathrm{o}})$ floating-point operations. The total operation count to solve (5) and (6) is approximately $\mathrm{O}[N_{\mathrm{o}}(N_{cg}n_{\mathrm{o}}^2 + n_{\mathrm{g}}n_{\mathrm{o}})]$, where N_{o} is the number of outer (quasi-Newton) iterations performed. In the current PSAS implementation, using conventional data only, $N_{\mathrm{cg}} \simeq 10$ and $n_{\mathrm{g}} \simeq 10n_{\mathrm{o}}$. Therefore approximately the same amount of computation is required for solving (5) as (6).

4. Assimilation of radiances

The assimilation of cloud-cleared radiances using a variational approach is operational at both the US National Centers for Environmental Prediction (Derber and Wu 1996) and the European Centre for Medium-Range Weather Forecasts (Andersson *et al.* 1994; Courtier *et al.* 1998; Rabier *et al.* 1998; Andersson *et al.* 1998). Assimilation of radiances involves utilizing radiance measurements **y** as the observation vector $\mathbf{w}^{\mathrm{o}}$ and specifying the radiance error covariance $\mathbf{R}^{y}$ as the observation error covariance $\mathbf{R}^{\mathrm{o}}$ in (1). In addition, an observation operator h must be specified that maps state variables to the radiances.

The observed radiance from a remote-sensing instrument can be expressed by

$$\mathbf{y} = f(\mathbf{z}^{\mathrm{t}}, \mathbf{b}^{\mathrm{t}}) + \delta f(\mathbf{z}^{\mathrm{t}}, \mathbf{b}^{\mathrm{t}}, \mathbf{b}'^{\mathrm{t}}) + \boldsymbol{\epsilon}^{\mathrm{measurement}}, \tag{7}$$

(Rodgers, personal communication): $\mathbf{z}^{\mathrm{t}}$ is the true atmospheric state at the observation location; f is an approximate radiative transfer or empirical model relating the true atmospheric state $\mathbf{z}^{\mathrm{t}}$ to radiance space using a vector of true parameters, $\mathbf{b}^{\mathrm{t}}$, such as spectral line

data and other calibration parameters; δf is the difference between the true observation operator and the approximate model f (e.g. resulting from discretization errors and errors in the physics of the approximate radiative-transfer model); $\mathbf{b}'^{\mathrm{t}}$ denotes parameters not included in f, and $\boldsymbol{\epsilon}^{\mathrm{measurement}}$ is the measurement error which includes detector noise, calibration errors, and any errors introduced by pre-processing of the data including cloud clearing, adjustment to nadir-viewing conditions, etc. The true state at the observation location, $\mathbf{z}^{\mathrm{t}}$, is given by

$$\mathbf{z}^{\mathrm{t}} = \mathcal{I}\mathbf{w}^{\mathrm{t}}, \tag{8}$$

where $\mathcal{I}$ is an interpolation operator relating a state on the analysis grid to the observation location. The radiative-transfer model is often linearized using a Taylor series expanded about an estimate of the state. For an iterative analysis or retrieval such as in (2), the linearization state is the state estimate at iteration i, denoted $\mathbf{z}^i$, i.e.

$$\begin{aligned} f(\mathbf{z}^{\mathrm{t}}) &= f(\mathbf{z}^i) + F_z^{\mathrm{t}}(\mathbf{z}^{\mathrm{t}} - \mathbf{z}^i) + \mathrm{O}(\mathbf{z}^{\mathrm{t}} - \mathbf{z}^i)^2 + \ldots \\ &= f(\mathbf{z}^i) + F_z^i(\mathbf{z}^{\mathrm{t}} - \mathbf{z}^i) + (F_z^{\mathrm{t}} - F_z^i)(\mathbf{z}^{\mathrm{t}} - \mathbf{z}^i) + \mathrm{O}(\mathbf{z}^{\mathrm{t}} - \mathbf{z}^i)^2 + \ldots \\ &= f(\mathbf{z}^i) + F_z^i(\mathbf{z}^{\mathrm{t}} - \mathbf{z}^i) + \boldsymbol{\epsilon}^{\mathrm{linear}}, \end{aligned} \tag{9}$$

where

$$F_z^i = \left.\frac{\partial f}{\partial \mathbf{z}}\right|_{\mathbf{z}=\mathbf{z}^i}, \tag{10}$$

and $\boldsymbol{\epsilon}^{\mathrm{linear}}$ will be referred to as the linearization error. The linearization error clearly depends upon the difference between the ith estimate of the state and the true state, and is likely to be correlated within certain spectral intervals.

To assimilate radiances correctly, an appropriate radiance error covariance $\mathbf{R}^y = \langle \boldsymbol{\epsilon}^y (\boldsymbol{\epsilon}^y)^{\mathrm{T}} \rangle$ must be specified. The total radiance error $\boldsymbol{\epsilon}^y$ is obtained by combining (7)–(10), i.e.

$$\boldsymbol{\epsilon}^y \equiv \mathbf{y} - \left(f(\mathbf{z}^i, \mathbf{b}) + F_z^i(\mathbf{z}^{\mathrm{t}} - \mathbf{z}^i) \right) = \boldsymbol{\epsilon}^{\mathrm{measurement}} + \frac{\partial f}{\partial b}(\mathbf{b}^{\mathrm{t}} - \mathbf{b}) + \delta f + \boldsymbol{\epsilon}^{\mathrm{linear}}, \tag{11}$$

where $\mathbf{b}$ are the parameters used in the observation operator. Although in some instances it may be possible to model $\boldsymbol{\epsilon}^y$, practical implementations will be likely to rely on statistical modelling from innovation sequences, or some form of on-line parameter estimation such as discussed in Dee (1995). Radiative-transfer parameter errors are usually associated with systematic errors (i.e. state-dependent errors that will be spatially correlated) as discussed by Eyre *et al.* (1993). The systematic errors may be corrected for in part by utilizing independent observations such as radiosondes, and/or forecasts or analyses that are assumed to be unbiased. This has been done, for example, with radiances from the TIROS* Operation Vertical Sounder (TOVS) (e.g. Eyre 1992; Susskind and Pfaendtner 1989). However, it is unlikely that the systematic errors can be completely eliminated. As a detailed treatment of systematic errors is beyond the scope of this paper, we treat radiative-transfer modelling errors as random and assume both background and observation errors to be unbiased, i.e. $\langle \boldsymbol{\epsilon} \rangle = 0$.

5. Assimilation of physical-space retrievals

Remotely sensed data have been traditionally assimilated in the form of physical-space retrievals. In this approach, radiances are used to produce familiar data types such as

* Television Infra-Red Observation Satellite.

temperature or humidity profiles that are used in a DAS. Specifying the retrieval $z \in \Re^{n_z}$ as the observation vector $\mathbf{w}^o$ in (1), the observation operator h is a linear interpolation operator so that the iterated form of (2) reduces to

$$\mathbf{w}^a = \mathbf{w}^f + \left(\mathbf{P}^f \mathcal{J}^T\right) \left(\mathcal{J}\mathbf{P}^f \mathcal{J}^T + \mathbf{R}^z\right)^{-1} \left(\mathbf{z} - \mathcal{J}\mathbf{w}^f\right), \tag{12}$$

where $\mathbf{R}^z = \langle \boldsymbol{\epsilon}^z (\boldsymbol{\epsilon}^z)^T \rangle$ is the retrieval error covariance, and $\mathcal{J} \in \Re^{n_z} \times \Re^{n_g}$ is the interpolation operator used above. A more general form of (12) that includes the retrieval-forecast-error cross-covariance, denoted by $\mathbf{X} = \langle \boldsymbol{\epsilon}^z (\boldsymbol{\epsilon}^f)^T \rangle$, is given by

$$\mathbf{w}^a = \mathbf{w}^f + \left(\mathbf{P}^f \mathcal{J}^T - \mathbf{X}^T\right) \left(\mathcal{J}\mathbf{P}^f \mathcal{J}^T + \mathbf{R}^z - \mathcal{J}\mathbf{X}^T - \mathbf{X}\mathcal{J}^T\right)^{-1} \left(\mathbf{z} - \mathcal{J}\mathbf{w}^f\right). \tag{13}$$

Operational implementations of retrieval assimilation often use retrieval error covariances that are statistically-derived, and assumed to be stationary and horizontally isotropic. In addition, the retrieval-forecast-error cross-covariance matrix $\mathbf{X}$ is often neglected as a result of the difficulty associated with modelling it and, more importantly, the potential for creating numerical instability as discussed by Eyre *et al.* (1993).

The retrieval of geophysical parameters $\mathbf{z}$ from radiance observations $\mathbf{y}$ is a nonlinear (in general) estimation process that can be represented by

$$\mathbf{z} = \mathbf{D}(\mathbf{y}, \mathbf{b}, \mathbf{z}^p) = \mathbf{D}(f[\mathbf{z}^t, \mathbf{b}^t] + \boldsymbol{\epsilon}^y, \mathbf{z}^p), \tag{14}$$

where $\mathbf{z}^p$ is a prior estimate of the state used in the retrieval algorithm, f is the radiative-transfer model encountered in the previous section, and $\boldsymbol{\epsilon}^y$ is the radiance error given by (11). Because f is nonlinear, (14) can linearized about an estimate of the state such as the prior estimate $\mathbf{z}^p$, i.e.

$$\mathbf{z} - \mathbf{z}^p = \left[\mathbf{D}(f(\mathbf{z}^p, \mathbf{b}), \mathbf{b}, \mathbf{z}^p) - \mathbf{z}^p\right] + \mathbf{D}_y \left[F_z \left(\mathbf{z}^t - \mathbf{z}^p\right) + \boldsymbol{\epsilon}^y\right], \tag{15}$$

(Rodgers 1990) where

$$\mathbf{D}_y = \left.\frac{\partial \mathbf{D}}{\partial y}\right|_{z=z^p}. \tag{16}$$

The first bracketed term on the r.h.s. of (15) is called the *transfer function bias*. For most physical retrievals, the algorithm returns the prior estimate when presented with noiseless data consistent with it (i.e. with $\mathbf{y} = f(\mathbf{z}^p, \mathbf{b})$). In such cases transfer function bias is insignificant and (15) reduces to

$$\mathbf{z} - \mathbf{z}^p = \mathbf{D}_y \left[\mathbf{y} - f(\mathbf{z}^p)\right] \tag{17}$$

(Eyre 1987). Then (15) or (17) can be rearranged to give the retrieval error $\boldsymbol{\epsilon}^z$, i.e.

$$\boldsymbol{\epsilon}^z = \mathbf{z} - \mathbf{z}^t = (\mathbf{I} - \mathbf{A})\, \boldsymbol{\epsilon}^p + \mathbf{D}_y \boldsymbol{\epsilon}^y \tag{18}$$

(e.g. Eyre 1987; Rodgers 1990) where $\mathbf{A} = \mathbf{D}_y F_z$ is called the averaging kernel (Backus and Gilbert 1970), and $\boldsymbol{\epsilon}^p = \mathbf{z}^p - \mathbf{z}^t$ is the prior-estimate error. The first term on the r.h.s. of (18) may be thought of as a *smoothing* error, which results from the use of prior information in components of the state space that the observing system is not able to measure accurately. The second term results from the propagation of radiance errors including all terms in (11).

(*a*) *Assimilation of interactive retrievals*

In the specific case of interactive retrievals, a forecast from a DAS interpolated to the observation location is used as the prior estimate (i.e. $\mathbf{z}^p = \mathcal{J}\mathbf{w}^f$). Then, using (18) we can derive an expression for the retrieval-forecast-error cross-covariance, i.e.

$$\mathbf{X} = \langle \boldsymbol{\epsilon}^z (\boldsymbol{\epsilon}^f)^T \rangle = (\mathbf{I} - \mathbf{A})\, \mathcal{J} \langle \boldsymbol{\epsilon}^f (\boldsymbol{\epsilon}^f)^T \rangle = (\mathbf{I} - \mathbf{A})\, \mathcal{J}\mathbf{P}^f, \tag{19}$$

where we have assumed that $\langle \boldsymbol{\epsilon}^y (\boldsymbol{\epsilon}^f)^T \rangle = 0$. Substitution of (19) into (13) gives

$$\mathbf{w}^a = \mathbf{w}^f + \mathbf{P}^f (\mathbf{A}\mathcal{J})^T \left[(\mathbf{A}\mathcal{J})\, \mathbf{P}^f\, (\mathbf{A}\mathcal{J})^T + \mathbf{D}_y \mathbf{R}^y \mathbf{D}_y^T \right]^{-1} (\mathbf{z} - \mathcal{J}\mathbf{w}^f)\,, \tag{20}$$

where $\mathbf{R}^y$ is the radiance observation error covariance encountered in section 4. This result was first derived by Ménard (personal communication) and shows that (13) can be simplified to eliminate terms involving $\mathbf{X}$. Then, (20) is of the same form as (12), where the operator $\mathbf{A}\mathcal{J}$ replaces the interpolation operator $\mathcal{J}$, and the retrieval error covariance is replaced by the propagation of radiance errors $\boldsymbol{\epsilon}^y$ to physical-space ($\mathbf{D}_y \mathbf{R}^y \mathbf{D}_y^T$). For any retrieval requiring the use of prior information, $\mathbf{A}\mathcal{J}$ is a stronger means of smoothing than the interpolation operator $\mathcal{J}$. It can be shown that this additional smoothing causes the innovation covariance to become singular when $\mathbf{A} \neq \mathbf{I}$. This conditioning problem, which is not unique to a PSAS, has been discussed by Eyre *et al.* (1993) and Purser (1990). Eyre *et al.* (1993) have adapted an approach given by Lorenc *et al.* (1986), which involves mapping 1D retrievals into a reduced space and then modifying both the retrievals and retrieval error variances, in order to overcome the numerical instability problem and to produce a more consistent assimilation of 1D retrievals for the case in which $\mathbf{A} \neq \mathbf{I}$. We discuss the trivial case of $\mathbf{A} = \mathbf{I}$ below.

(*b*) *Assimilation of retrievals with no prior information*

If little or no prior information is used in the retrieval, then by definition $\mathbf{A} \approx \mathbf{I}$ and therefore $\mathbf{X} = (\mathbf{I} - \mathbf{A})\mathcal{J}\langle \boldsymbol{\epsilon}^p (\boldsymbol{\epsilon}^f)^T \rangle \approx 0$. In this case the retrieval may be consistently assimilated by substituting $\mathbf{R}^z = \mathbf{D}_y \mathbf{R}^y \mathbf{D}_y^T$ into (12). This approach has been used to assimilate constituent data from the Cryogenic Limb Array Etalon Spectrometer. In the next section, we describe methods to filter or remove prior information from retrievals that have non-trivial averaging kernels ($\mathbf{A} \neq \mathbf{I}$) so that $\mathbf{X} \simeq 0$ and (12) can be used.

6. Assimilation of phase-space retrievals

Consider a new data type $\boldsymbol{\alpha}$ derived from retrievals by a linear transformation, e.g.

$$\boldsymbol{\alpha} = \mathbf{U}_R^T \mathbf{z}, \tag{21}$$

where $\mathbf{U}_R \in \Re^{n_z} \times \Re^{n_\ell}$ ($n_\ell \leqslant n_z$) reduces the dimension of the retrieval vector. Because the cost of the conjugate gradient solver in a PSAS approximately scales as the square of the number of observations, a linear transformation that significantly reduces the size of the observation vector will reduce the cost of the analysis. The statistical analysis equation in terms of the new data type $\boldsymbol{\alpha}$ becomes

$$\begin{aligned} \mathbf{w}^a &= \mathbf{w}^f + \mathbf{P}^f \mathcal{J}^T \mathbf{U}_R \left(\mathbf{U}_R^T \mathcal{J} \mathbf{P}^f \mathcal{J}^T \mathbf{U}_R + \mathbf{R}^\alpha \right)^{-1} \left(\boldsymbol{\alpha} - \mathbf{U}_R^T \mathcal{J} \mathbf{w}_f \right) \\ &= \mathbf{w}^f + \mathbf{K}^\alpha \left(\boldsymbol{\alpha} - \mathbf{U}_R^T \mathcal{J} \mathbf{w}_f \right)\,, \end{aligned} \tag{22}$$

where

$$\mathbf{R}^{\alpha} = \langle \boldsymbol{\epsilon}^{\alpha} (\boldsymbol{\epsilon}^{\alpha})^{\mathrm{T}} \rangle = \mathbf{U}_R^{\mathrm{T}} \mathbf{R}^{z} \mathbf{U}_R \,, \tag{23}$$

is the error covariance of the new data type $\boldsymbol{\alpha}$, and $\boldsymbol{\epsilon}^{\alpha} = \boldsymbol{\alpha} - \mathbf{U}_R^{\mathrm{T}} \mathbf{z}^{\mathrm{t}}$. We have neglected the retrieval-forecast-error cross-covariance $\mathbf{X}$. In the next two sections, we present several ways to design a compact linear operator $\mathbf{U}_R$ to ensure that $\mathbf{X} \approx 0$.

(*a*) *Assimilation of null-space filtered (NSF) retrievals*

(i) *Method 1.* In order to clarify the relationship between the retrieval, the true state, and the prior estimate, (18) can be rearranged to give

$$\mathbf{z} = \mathbf{A}\mathbf{z}^{\mathrm{t}} + (\mathbf{I} - \mathbf{A})\mathbf{z}^{\mathrm{p}} + \mathbf{D}_y \boldsymbol{\epsilon}^{y} \tag{24}$$

(Rodgers, personal communication). The 1D retrieval $\mathbf{z} \in \Re^{n_z}$ is partially made up of a smoothed version of the true atmospheric state ($\mathbf{A}\mathbf{z}^{\mathrm{t}}$). The details of the state that cannot be resolved by the instrument are supplied by prior information via $(\mathbf{I} - \mathbf{A})\mathbf{z}^{\mathrm{p}}$, with added error $\mathbf{D}_y \boldsymbol{\epsilon}^{y}$ resulting from the propagation of radiance errors. In order to simplify the error characterization for data assimilation, we seek a linear transformation of the retrieval that eliminates (or renders negligibly small) the contribution from the prior estimate $\mathbf{z}^{\mathrm{p}}$. One method to accomplish this is to project the retrieval $\mathbf{z}$ onto the null-space of the operator $\mathbf{I} - \mathbf{A}$. Specifically, consider the SVD

$$\mathbf{I} - \mathbf{A} = \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^{\mathrm{T}}, \tag{25}$$

where $\mathbf{U}_1$ and $\mathbf{V}_1$ are matrices of left and right singular vectors of $\mathbf{I} - \mathbf{A}$ (the subscript 1 denotes method 1), and $\mathbf{S}_1$ is a diagonal matrix of singular values sorted in descending order. Let $\mathbf{U}_1$ be expressed as the concatenation of leading and trailing singular vectors, $\mathbf{U}_{1\mathrm{L}}$ and $\mathbf{U}_{1\mathrm{T}}$, respectively, i.e. $\mathbf{U}_1 = (\mathbf{U}_{1\mathrm{L}} | \mathbf{U}_{1\mathrm{T}})$. We take $\mathbf{U}_{1\mathrm{T}} \in \Re^{n_z} \times \Re^{n_\ell}$ to correspond to negligible singular values $\mathbf{S}_{1\mathrm{T}} \in \Re^{n_\ell} \times \Re^{n_\ell}$, so that

$$(\mathbf{U}_{1\mathrm{T}})^{\mathrm{T}} (\mathbf{I} - \mathbf{A})\mathbf{z}^{\mathrm{p}} = \mathbf{S}_{1\mathrm{T}} (\mathbf{V}_{1\mathrm{T}})^{\mathrm{T}} \mathbf{z}^{\mathrm{p}} \approx 0. \tag{26}$$

We will refer to the subspace spanned by the leading modes, $\mathbf{U}_{1\mathrm{L}}$, as the null-space of the observations. Although not strictly a null-space in the mathematical sense, the observation null-space can be thought of as components of the state space that the instrument cannot measure accurately. A new data type $\boldsymbol{\alpha}_1 \equiv (\mathbf{U}_{1\mathrm{T}})^{\mathrm{T}} \mathbf{z}$ has a corresponding error covariance $\mathbf{R}^{\alpha_1}$ with negligible contribution from the prior information, i.e. rearranging (23), (24), and (26)

$$\mathbf{R}^{\alpha_1} \equiv (\mathbf{U}_{1\mathrm{T}})^{\mathrm{T}} \mathbf{R}^{z} \mathbf{U}_{1\mathrm{T}} \approx (\mathbf{U}_{1\mathrm{T}})^{\mathrm{T}} \mathbf{D}_y \mathbf{R}^{y} \mathbf{D}_y^{\mathrm{T}} \mathbf{U}_{1\mathrm{T}}. \tag{27}$$

The data type $\boldsymbol{\alpha}_1$, a phase-space retrieval, is consistently assimilated using $\mathbf{U}_{\mathbf{R}} = \mathbf{U}_{1\mathrm{T}}$ in (22).

(ii) *Method 2.* Another approach to eliminate prior information from retrievals involves projecting the retrieval onto the null-space of the smoothing error $(\mathbf{A} - \mathbf{I})\boldsymbol{\epsilon}^{\mathrm{p}}$ from (18). The smoothing error covariance, denoted $\mathbf{R}^{s}$, and its eigen-decomposition are given by

$$\mathbf{R}^{s} = (\mathbf{A} - \mathbf{I})\, \mathbf{P}^{\mathrm{p}}\, (\mathbf{A} - \mathbf{I})^{\mathrm{T}} = \mathbf{U}_2 \Phi \mathbf{U}_2^{\mathrm{T}}, \tag{28}$$

where Φ is a diagonal matrix of real, non-negative eigenvalues. $\mathbf{U}_2$ can be expressed as $\mathbf{U}_2 \equiv [\mathbf{U}_{2\mathrm{L}} \mid \mathbf{U}_{2\mathrm{T}}]$, where $\mathbf{U}_{2\mathrm{L}}$ and $\mathbf{U}_{2\mathrm{T}}$ are leading and trailing modes (eigenvalues are sorted in descending order). We now show that a new data type $\boldsymbol{\alpha}_2 \equiv (\mathbf{U}_{2\mathrm{T}})^{\mathrm{T}} \mathbf{z}$ has an error

$\boldsymbol{\epsilon}^{\alpha_2} = \boldsymbol{\alpha}_2 - (\mathbf{U}_{2\mathrm{T}})^{\mathrm{T}}\mathbf{z}^{\mathrm{t}}$ with negligible contribution from the prior estimate. Using (18), (23), and (28), $\mathbf{R}^{\alpha_2}$ is given by

$$\begin{aligned}\mathbf{R}^{\alpha_2} &= (\mathbf{U}_{2\mathrm{T}})^{\mathrm{T}}\mathbf{U}_2\Phi\mathbf{U}_2^{\mathrm{T}}\mathbf{U}_{2\mathrm{T}} + (\mathbf{U}_{2\mathrm{T}})^{\mathrm{T}}\mathbf{D}_y\mathbf{R}^y\mathbf{D}_y^{\mathrm{T}}\mathbf{U}_{2\mathrm{T}} \\ &= \Phi_{\mathrm{T}} + (\mathbf{U}_{2\mathrm{T}})^{\mathrm{T}}\mathbf{D}_y\mathbf{R}^y\mathbf{D}_y^{\mathrm{T}}\mathbf{U}_{2\mathrm{T}}, \end{aligned} \tag{29}$$

where Φ_{T} is a diagonal matrix of trailing eigenvalues. Assuming the trailing eigenvalues to be small, i.e. $||\Phi_{\mathrm{T}}|| \ll ||\mathbf{U}_{2\mathrm{T}}^{\mathrm{T}}\mathbf{D}_y\mathbf{R}^y\mathbf{D}_y^{\mathrm{T}}\mathbf{U}_{2\mathrm{T}}||$, (29) reduces to

$$\mathbf{R}^{\alpha_2} \simeq (\mathbf{U}_{2\mathrm{T}})^{\mathrm{T}}\mathbf{D}_y\mathbf{R}^y\mathbf{D}_y^{\mathrm{T}}\mathbf{U}_{2\mathrm{T}} = (\mathbf{U}_{2\mathrm{T}})^{\mathrm{T}}\mathbf{R}^z\mathbf{U}_{2\mathrm{T}}. \tag{30}$$

Again, the data type $\boldsymbol{\alpha}_2$ is a phase-space retrieval that can be assimilated using (22) with $\mathbf{U}_R = \mathbf{U}_{2\mathrm{T}}$.

(iii) *Implementation issues.* The above methods can be thought of as observation null-space filters (NSFs). NSF retrieval assimilation has several advantages over physical-space retrieval assimilation. In addition to removing prior information from the retrieval, eliminating modes of the decomposition results in data compression that should reduce the amount of computation required in a PSAS. If the state dependence of averaging-kernel $\mathbf{A}$ is small, a constant or small set of fixed trailing eigenvectors or singular vectors may be defined prior to assimilation, further reducing computation in the DAS. Another potential advantage is that radiances are not required so that implementation may be simplified, e.g. if the radiances are not available. These methods do, however, optimally require knowledge about how the retrieval is performed; in some cases this information may not be generally available. Finally, when the retrieval is severely constrained (i.e. when a significant amount of prior information is contained in the sub-space spanned by trailing modes $\mathbf{U}_{\mathrm{T}}$), these methods may not be as accurate as radiance assimilation. This is because the null-space filter may not completely eliminate all of the prior information, or it may remove some of the useful information contained in the radiances. A more optimal approach requiring radiances is given in the following subsection.

(*b*) *Assimilation of partial eigen-decomposition (PED) retrievals*

(i) *PED retrieval.* In the previous subsection, operators were designed to filter prior information from physical-space retrievals to yield a modified retrieval in phase-space. We now define a phase-space retrieval process that completely eliminates prior information. Recall that a retrieval can be represented as a nonlinear estimation process $\mathbf{D}(f, \mathbf{z}^{\mathrm{p}}...)$ with associated operators $\mathbf{D}_y = \partial\mathbf{D}/\partial\mathbf{y}$ and $\mathbf{F}_z = \partial f/\partial z$. For convenience, we will consider the case of $\mathbf{D}$ formulated as an unconstrained weighted least-squares retrieval. $\mathbf{D}$ is linearized about a prior estimate $\mathbf{z}^{\mathrm{p}}$, i.e.

$$\mathbf{z} = \mathbf{z}^{\mathrm{p}} + [\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z]^{-1}\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}[\mathbf{y} - f(\mathbf{z}^{\mathrm{p}}, \mathbf{b})]. \tag{31}$$

The unconstrained solution is obtained by setting the prior-estimate error equal to infinity (inverse of the prior-estimate error covariance equal to zero) in the 1D version of (2). Then, $\mathbf{D}_y$ is given by

$$\mathbf{D}_y = \left(\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z\right)^{-1}\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}, \tag{32}$$

and $\mathbf{A} = \mathbf{D}_y\mathbf{F}_z = \mathbf{I}$. Substituting (32) into (18), the retrieval error covariance, $\mathbf{R}^z = \langle\boldsymbol{\epsilon}^z(\boldsymbol{\epsilon}^z)^{\mathrm{T}}\rangle$, is given by

$$\mathbf{R}^z = \mathbf{D}_y\mathbf{R}^y\mathbf{D}_y^{\mathrm{T}} = \left(\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z\right)^{-1} \tag{33}$$

and has no contribution from the prior estimate. Evaluating (31)–(33) is straightforward when $\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z$ is nonsingular. However, it is often the case that $[\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z]$ is singular or nearly singular, so that either the solution to (31)–(33) does not exist, or it exists but the propagation of radiance errors produces large errors in some components of the retrieval.

Alternatively, we can formulate an unconstrained phase-space retrieval in terms of coefficients of the well-determined eigenvectors of $[\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z]$ (i.e. coefficients with small errors). Consider the eigen-decomposition of $[\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z]$, i.e.

$$[(\mathbf{F}_z)^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z] = (\mathbf{R}^z)^{-1} = \mathbf{U}_3\boldsymbol{\Lambda}\mathbf{U}_3^{\mathrm{T}}, \tag{34}$$

where eigenvalues are sorted in descending order. This decomposition is known as the Hotelling, discrete Karhunen–Loeve, eigenvector, or principal component transform (Hotelling 1933; Karhunen 1947; Loeve 1948). As before, the columns of $\mathbf{U}_3$ (i.e. the eigenvectors of $(\mathbf{R}^z)^{-1}$ or $\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z$) can be written as $\mathbf{U}_3 = [\mathbf{U}_{3\mathrm{L}} \mid \mathbf{U}_{3\mathrm{T}}]$, where $\mathbf{U}_{3\mathrm{L}}$ and $\mathbf{U}_{3\mathrm{T}}$ are the leading and trailing eigenvectors or modes and $\mathbf{U}_{3\mathrm{L}} \in \Re^{n_z} \times \Re^{n_\ell}$. Eigenvectors in the null-space of $(\mathbf{R}^z)^{-1}$ will have zero eigenvalues, and modes that are not well-determined by the measurements (i.e. the trailing modes) will have small eigenvalues or large errors. A new data type $\boldsymbol{\alpha}_3 \equiv \mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{z}$ can be retrieved by linearizing the operators about an initial state $\mathbf{z}^{\mathrm{p}}$, i.e.

$$\boldsymbol{\alpha}_3 = \mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{z}^{\mathrm{p}} + \mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{U}_3\boldsymbol{\Lambda}^{-1}\mathbf{U}_3^{\mathrm{T}}\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}[\mathbf{y} - f(\mathbf{z}^{\mathrm{p}}, \mathbf{b})]. \tag{35}$$

Because $\mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{U}_3$ is the rectangular matrix $[\mathbf{I}_\ell \mid 0]$, it follows that

$$\mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{U}_3\boldsymbol{\Lambda}^{-1}\mathbf{U}_3^{\mathrm{T}} = \boldsymbol{\Lambda}_L^{-1}\mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}. \tag{36}$$

Substituting (36) in (35) gives

$$\boldsymbol{\alpha}_3 = \mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{z}^{\mathrm{p}} + \boldsymbol{\Lambda}_L^{-1}\mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}[\mathbf{y} - f(\mathbf{z}^{\mathrm{p}}, \mathbf{b})]. \tag{37}$$

Then, substituting $\mathbf{D}_y$ given by

$$\mathbf{D}_y = \boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1} \tag{38}$$

into (18) and rearranging, we can express $\mathbf{R}^{\alpha_3}$ as

$$\mathbf{R}^{\alpha_3} = \boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{U}_{3\mathrm{L}}^{\mathrm{T}}\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{R}^y[(\mathbf{R}^y)^{-1}]^{\mathrm{T}}\mathbf{F}_z\mathbf{U}_{3\mathrm{L}}(\boldsymbol{\Lambda}_{\mathrm{L}}^{-1})^{\mathrm{T}} = \boldsymbol{\Lambda}_{\mathrm{L}}^{-1}, \tag{39}$$

so that $\mathbf{R}^{\alpha_3}$ has no contribution from the prior estimate. The prior estimate is used only as a linearization point. An equivalent form of the PED solution may be derived by the SVD of $(\mathbf{R}^y)^{-1/2}\mathbf{F}_z$, i.e. $(\mathbf{R}^y)^{-1/2}\mathbf{F}_z = \widehat{\mathbf{U}}_3\mathbf{D}\widehat{\mathbf{V}}_3^{\mathrm{T}}$, where the singular values, $\mathbf{D}$, are the square roots of the eigenvalues in $\boldsymbol{\Lambda}$, and the columns of $\widehat{\mathbf{V}}_3$ are the same as the columns of $\mathbf{U}_3$ in the above equations (Rodgers, personal communication). The PED transformation minimizes the truncation error (Koschman 1954) which is defined as the mean-square difference between the exact and truncated-mode solution. The error introduced by truncating the last $n_z - n_\ell$ modes, denoted $\boldsymbol{\epsilon}^{\mathrm{trunc}}$, is given by

$$\boldsymbol{\epsilon}^{\mathrm{trunc}} = \sum_{l=n_\ell+1}^{n_z} \boldsymbol{\Lambda}_{ll}. \tag{40}$$

As before, the data type $\boldsymbol{\alpha}_3$ can be assimilated in phase-space using (22) with $\mathbf{U}_R = \mathbf{U}_{3\mathrm{L}}$.

(ii) *Implementation issues.* The PED transformation operator, $\mathbf{U}_{3L}$, is optimal in terms of packing the information content of the radiances into leading modes. This optimal transformation is state-dependent as a result of the nonlinear radiative-transfer operator. In order to simplify the implementation of the PED retrieval assimilation, a fixed transformation $\mathbf{U}_{LF}$ (that does not necessarily diagonalize the retrieval error covariance) may be used in place of the optimal state-dependent transformation $\mathbf{U}_{3L}$. The truncation error of the suboptimal fixed PED transformation will be slightly larger than that of the optimal PED transformation for a given number of truncated modes. The use of a fixed transformation would more easily facilitate some form of spatial averaging of data prior to assimilation (i.e. *super-obbing* as described by Lorenc (1981)) that reduces the amount of computation required by a DAS such as a PSAS. If the radiative-transfer model is nearly linear, the off-diagonal elements of the retrieval error covariance may be negligible, so that one can find a fixed transformation that approximately diagonalizes the retrieval error covariance.

We can also project the unconstrained retrieval onto modes of the partial eigendecomposition of other operators that include the forecast-error covariance, such as the averaging kernel or the normalized precision matrix discussed by Huang and Purser (1996). These operators, which are normalized with respect to the forecast, may pack the information more efficiently in terms of its impact on the analysis. It will also be easier to interpret the eigenvalues of these operators, especially when state variables are of mixed units.

(*c*) *Implementation issues common to NSF and PED retrievals*

(i) *Truncation criteria.* In the case of either NSF or PED retrieval assimilation, the number of modes to preserve and assimilate must be determined such that the truncation error is acceptable. In addition, for NSF retrievals it should be verified that the amount of prior information in a given mode is negligible, and that useful information is not being discarded. It is possible to quantitatively assess the impact of a particular mode on a DAS, and the effect on the DAS of truncating a given number of modes. There are several measures of information content and data impact that have been used previously in remote-sensing applications (see e.g. Eyre 1990; Huang and Purser 1996; Prunet *et al.* 1998). We have used here two different methods of examining data impact for the PED retrieval assimilation example. On a mode-by-mode basis, one measure of data impact is the ratio of the analysis error variance to the forecast-error variance. This quantity will be scalar invariant, meaning that it is a quantity independent of the units or coordinates used, as defined by Purser and Huang (1993). The forecast-error variance for mode ℓ, p_ℓ^{f} (a scalar), is given by

$$p_\ell^{\mathrm{f}} \equiv \mathbf{u}_\ell^{\mathrm{T}} \mathcal{J} \mathbf{P}^{\mathrm{f}} \mathcal{J}^{\mathrm{T}} \mathbf{u}_\ell, \tag{41}$$

where $\mathbf{u}_\ell$ is the ℓth eigenvector. Similarly, the analysis error for mode ℓ, p_ℓ^{a}, is given by

$$p_\ell^{\mathrm{a}} \equiv \mathbf{u}_\ell^{\mathrm{T}} \mathcal{J} \mathbf{P}^{\mathrm{a}} \mathcal{J}^{\mathrm{T}} \mathbf{u}_\ell. \tag{42}$$

If the ratio $p_\ell^{\mathrm{a}}/p_\ell^{\mathrm{f}}$ is close to unity, the assimilation of the data type does not significantly reduce the analysis error for that mode, and therefore may be truncated without significant degradation to the DAS. For PED retrievals, substituting $\mathbf{P}^{\mathrm{a}} \equiv (\mathbf{I} - \mathbf{K}^{\alpha}\mathbf{H})\mathbf{P}^{\mathrm{f}}$, where $\mathbf{K}^{\alpha}$ is given by (22) and $\mathbf{H}$ is given by $\mathbf{U}_R^{\mathrm{T}}\mathcal{J}$, into (42), it can be shown that

$$\begin{aligned} p_\ell^{\mathrm{a}}/p_\ell^{\mathrm{f}} &= 1 - (\mathbf{u}_\ell^{\mathrm{T}} \mathcal{J} \mathbf{P}^{\mathrm{f}} \mathcal{J}^{\mathrm{T}} \mathbf{u}_\ell + \mathbf{\Lambda}_{ll}^{-1})^{-1} \mathbf{u}_\ell^{\mathrm{T}} \mathcal{J} \mathbf{P}^{\mathrm{f}} \mathcal{J}^{\mathrm{T}} \mathbf{u}_\ell \\ &= 1 - (p_\ell^{\mathrm{f}} + \mathbf{\Lambda}_{ll}^{-1})^{-1} p_\ell^{\mathrm{f}} = (1 + p_\ell^{\mathrm{f}} \mathbf{\Lambda}_{ll})^{-1}. \end{aligned} \tag{43}$$

Alternatively, another quantity to examine is the ratio of the vertically averaged analysis error variance, trace[$\mathbf{P}^{\mathrm{a}}$], to the vertically averaged forecast-error variance, trace[$\mathbf{P}^{\mathrm{f}}$],

for a given number of included modes. This ratio will indicate the reduction in total analysis error variance resulting from the assimilation of a particular mode or set of modes. This measure has the disadvantage that it is not invariant. However, it provides a convenient way to express data impact for the particular example used below.

(ii) *Linearization Issues.* The new data type $\boldsymbol{\alpha}$ and corresponding observation operator $\mathbf{U}_R$ in (22) are obtained in the above examples by linearizing about a prior estimate. For the PED retrieval, we may use either a forecast, a 1D retrieval, or a 3D analysis as a linearization point. Linearization error may be significant when the radiative-transfer operator is highly nonlinear (e.g. in the case of humidity) and when a forecast is used as the linearization point. The linearization error may be significantly reduced by performing a 1D retrieval prior to the eigen-decomposition. Linearizing in 1D, as in NSF retrieval assimilation, may be far less costly than linearizing in 3D as in some implementations of radiance assimilation.

7. Comparison of data assimilation methods using a 1D simulation

(a) *Experimental design*

The objective of this experiment is to compare assimilation methods in a limited 1D DAS using a Monte Carlo approach. For the experiment, the true state (truth), the analysed state (analysis), and the forecast state (the prior estimate) are an ensemble of temperature (geopotential height or thickness) profiles. Radiances from infrared-temperature sounding instruments are simulated, and the data are assimilated in 1D using several different methods to produce an analysis (i.e. horizontal correlations are ignored). The methods included are 1D radiance assimilation (equivalent to a 1D minimum variance interactive retrieval) and 1D assimilation of PED and NSF retrievals retaining different numbers of modes. The PED retrievals use the final 1D minimum variance retrieval as a linearization point, and the null-space filtering is also performed on the final 1D minimum variance retrieval. This experiment provides a comparison between data assimilation methods in a controlled environment, and also provides a rough comparison of the computation that will be required by a PSAS for the different methods in a full 3D implementation. Similar results should also be obtained in a 3D system provided that the linearization error is small in the 1D case.

Figure 1 shows three model-generated temperature profiles that provide a reasonable range of variability. A comparison of the different data assimilation methods is performed separately for the two extreme profiles (high- and low-latitude profiles) to show how the magnitude of the errors varies with state. The truth in each case consists of 1000 identical profiles. The forecast is produced by perturbing the true state according to the statistics of the forecast-error covariance. The forecast-error covariance ($\mathbf{P}^f$) used here is that for a 6-hour forecast from the 18-level (0.4, 1, 2, 5, 10, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, and 1000 mb) Goddard Earth Observing System (GEOS) 1.0 general circulation model. $\mathbf{P}^f$ was estimated from innovation (observed minus forecast) height time series over North America (da Silva *et al.* 1996).

Radiances (brightness temperatures) are simulated from the truth using a fast radiative-transfer algorithm with parametrizations similar to those used by Susskind *et al.* (1983). Errors are added to the channel brightness temperatures; they have a normal distribution and there is no correlation between spectral elements (channels). The assumed brightness-temperature error covariance, $\mathbf{R}^y$, is taken to be a diagonal matrix (channel-independent errors) where the elements are set equal to the specified errors, but with the addition of $(0.05\ \mathrm{K})^2$ to account for linearization error.

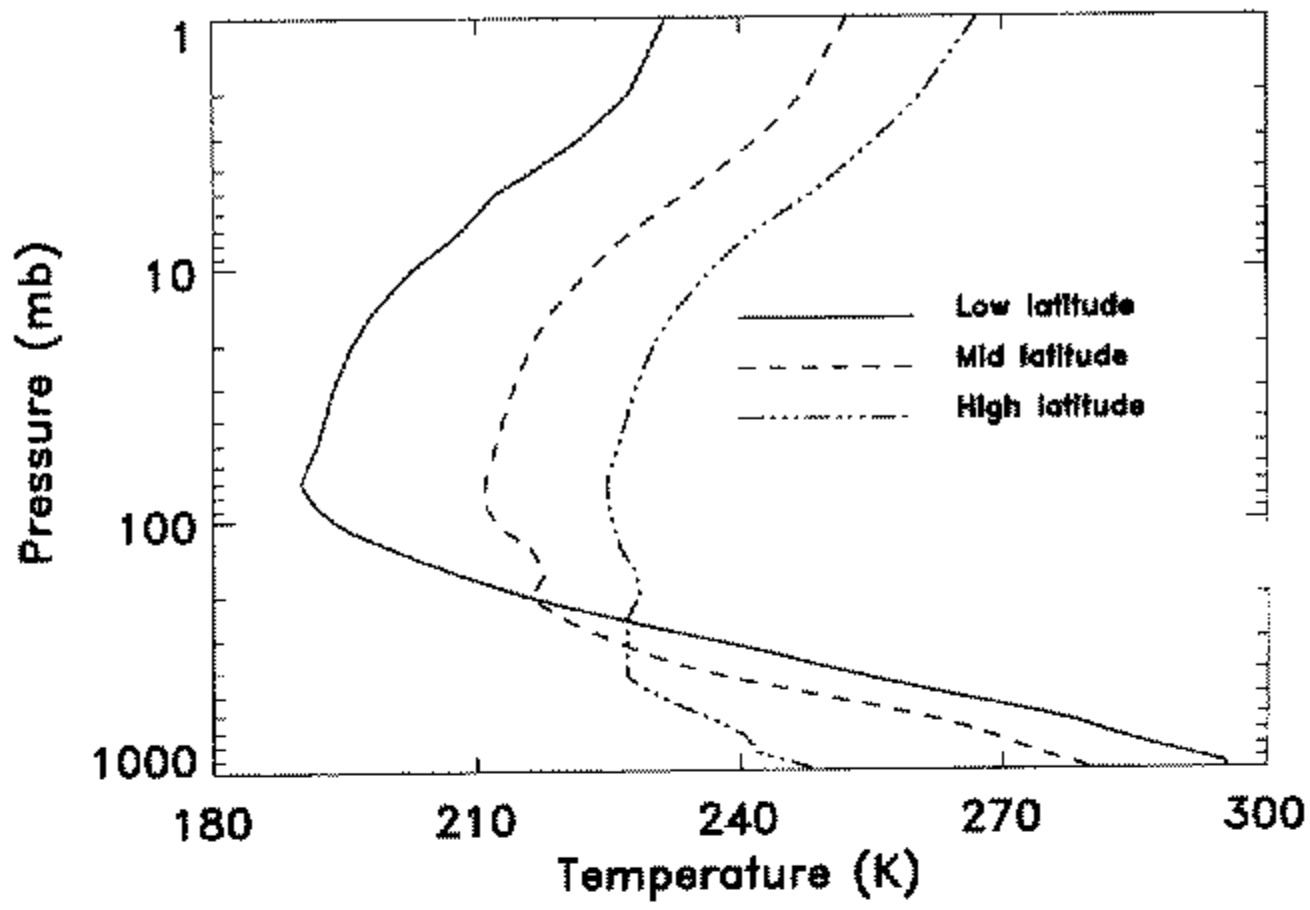

Figure 1. Model-generated temperature profiles used in the simulation.

For simplicity, we simulated clear-sky night-time conditions (i.e. no reflected solar radiation) and nadir-viewing conditions. The water vapour profile and surface parameters are assumed to be known; when these parameters are estimated, retrieval errors will in general be larger than those shown here. Another simplifying assumption is the exclusion of channel-correlated errors here. Channel-correlated errors will occur as a result of radiative-transfer modelling errors, and will ultimately limit the information content of the measurements. Although a number of simplifying assumptions are made here, a meaningful comparison between different assimilation approaches can be obtained because they are the same assumptions in all cases.

(*b*) *Instrument description*

We simulated radiances for two different infrared sounders: the Atmospheric Infrared Sounder (AIRS) and the High-resolution Infrared Sounder 2 (HIRS2). AIRS will fly on the NASA Earth Observing System afternoon (EOS-PM) platform in the year 2000. AIRS is a high-spectral-resolution grating spectrometer. The resolving power of AIRS, defined as the ratio of the frequency ν of a spectral element to its full-width half maximum bandwidth $\Delta\nu$, is approximately equal to 1200 over its full spectral range. AIRS has approximately 2600 contiguous spectral elements with some gaps, sampled twice per half width, that cover wavelength regions from 3.4 to 15.4 μm (650 to 2670 cm^{-1}). The spatial resolution of AIRS is approximately 13 km at nadir in the 705 km orbit of EOS-PM. AIRS will scan laterally, with satellite zenith angles ranging between $\pm 49.5°$. The specified single spot equivalent noise temperature for a given spectral-resolution element is 0.2 K at a 250 K scene for channels $\nu > 750\ \mathrm{cm}^{-1}$ and 0.35 K at a 250 K scene for channels at lower frequencies.

HIRS2 is one of 3 instruments contained in the TOVS package and is described in detail by Smith *et al.* (1979). HIRS2 first flew on TIROS-N launched in 1978 and has now flown for 20 years on NOAA operational satellites 6–14. HIRS2 has 19 infrared channels with a resolving power ($\nu/\Delta\nu$) of about 100. HIRS2 has approximately the same spatial resolution, spatial coverage, and signal-to-noise (per channel) as AIRS.

We have included in the simulation a total of 550 AIRS channels consisting of all those available between 650 and 742 cm^{-1}, between 2160 and 2270 cm^{-1}, and between

TABLE 1. CENTRAL WAVENUMBER ν AND SINGLE SPOT BRIGHTNESS TEMPERATURE ERRORS USED IN SIMULATION FOR HIRS2 CHANNELS AND AIRS CHANNELS AT THE SAME FREQUENCY.

HIRS channel	ν cm^{-1}	HIRS K	AIRS K
1	668.9	2.460	0.175
2	679.4	0.550	0.175
3	689.6	0.411	0.175
4	703.6	0.255	0.175
5	714.5	0.174	0.175
6	732.3	0.200	0.175
7	749.6	0.168	0.100
13	2191.3	0.281	0.100
14	2207.4	0.153	0.100
15	2236.4	0.224	0.100
16	2268.1	0.116	0.100

2379 and 2407 cm^{-1}. We used a total of 11 HIRS channels covering the first 2 AIRS wavelength regions (channels 1–7 and 13–16). Table 1 gives the central frequency and specified radiometric noise equivalent temperature ($NE\Delta T$) at a 250 K scene for HIRS2 channels. The specified $NE\Delta T$s are sometimes significantly higher than actual instrument performance, but do not include contributions to radiance errors from radiative-transfer modelling errors, calibration errors, pre-processing errors, etc. Here we use the specified $NE\Delta T$ values as an estimate of the total radiance error as they are comparable with radiance error estimates used elsewhere (e.g. Eyre *et al.* 1993). Table 1 also gives expected $NE\Delta T$s for single AIRS channels at the HIRS channel central frequencies. The AIRS $NE\Delta T$s used in the simulation were 0.1 K for channels $\nu > 738.55$ cm^{-1} and 0.175 K at a 250 K scene for channels at lower frequencies. These are the values used in simulation studies by the AIRS science team (H. H. Aumann, personal communication). We could have used all AIRS or HIRS channels and included other state variables (i.e. humidity, surface temperature and ozone) in the analysis as has been done in previous work (Eyre 1990; Huang and Purser 1996; Prunet *et al.* 1998). However, we have chosen this simple example for demonstration purposes.

Figures 2 and 3 show the Jacobian or channel temperature sensitivities for selected AIRS and HIRS channels, respectively, using the midlatitude temperature profile from above. Each curve represents $F_z = \partial\Theta_k/\partial T$ as a function of pressure or altitude, where Θ_k is the channel, k the brightness temperature, and T is the atmospheric temperature profile.

(c) *Radiance assimilation: linear error analysis*

Figure 4 shows predicted thickness errors (in the layers bounded by the 18 GEOS levels given above) for radiance assimilation computed using the linear error estimation theory described above. Also shown are specified forecast errors. The largest impact from the sounding data is in the stratosphere and lower troposphere, and the impact from AIRS is significantly greater than that from HIRS2. Figure 5 shows predicted thickness errors for both the low- and high-latitude cases using AIRS channels to illustrate the state dependence of the retrieval errors. The forecast-error covariance was generated in the northern hemisphere over land where conventional data is dense. In the southern hemisphere, where forecast errors are typically larger, the impact of AIRS and HIRS2 data will in general be

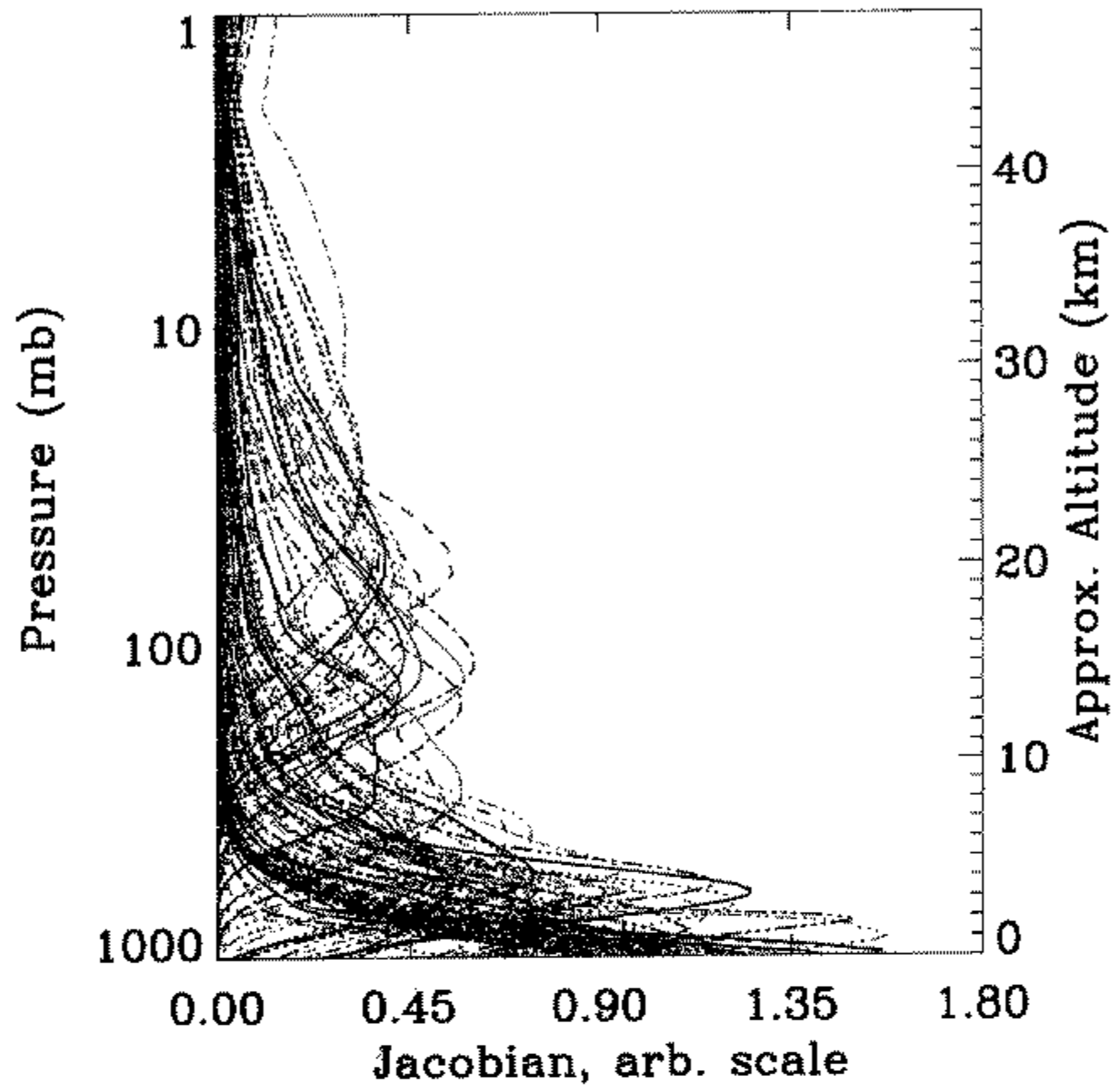

Figure 2. Jacobian or channel brightness temperature sensitivities to atmospheric temperature for selected Atmospheric Infrared Sounder (AIRS) channels (each curve is for one AIRS channel).

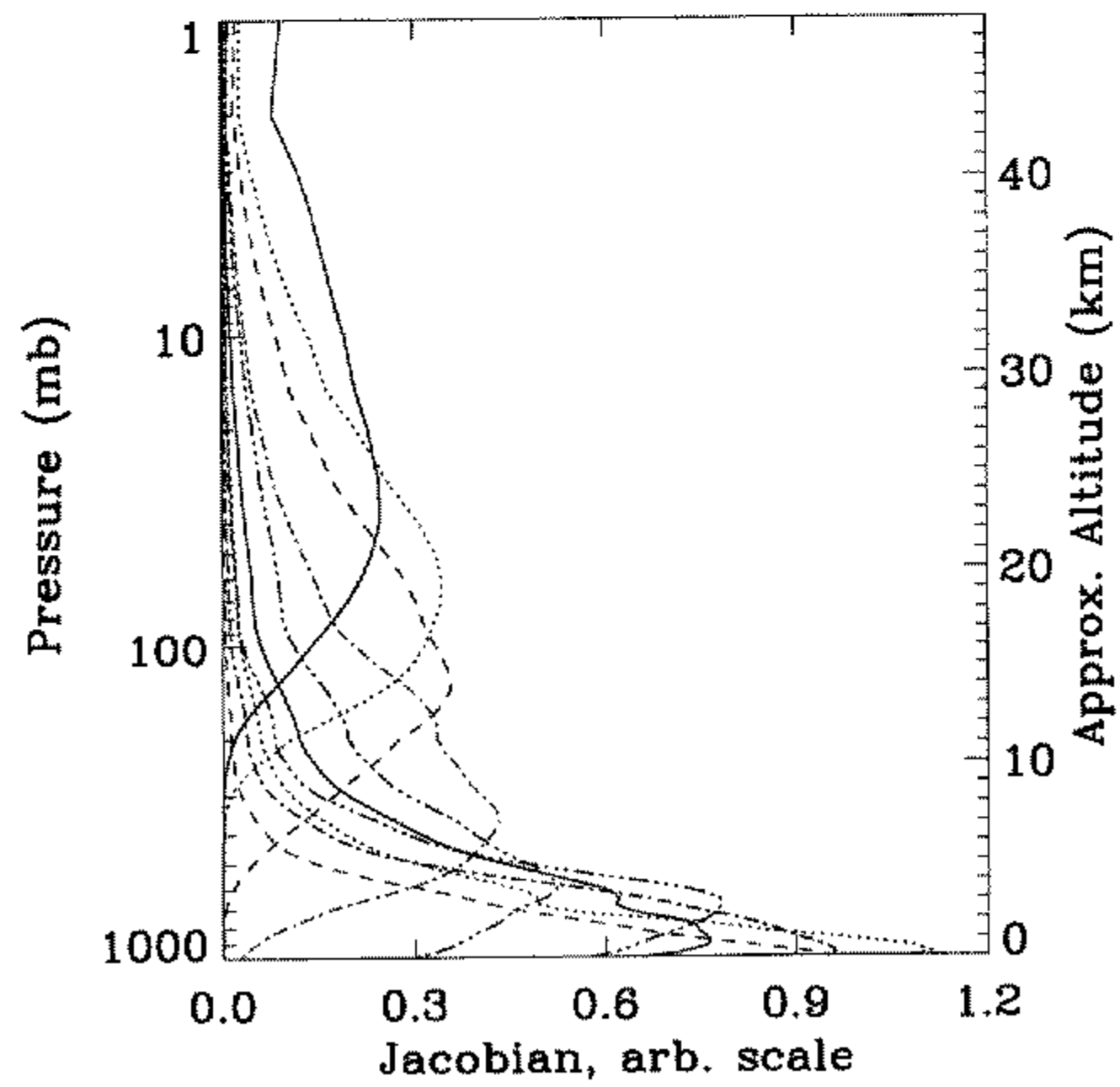

Figure 3. As Fig. 2 but for selected High-resolution Infrared Sounder 2 (HIRS2) channels.

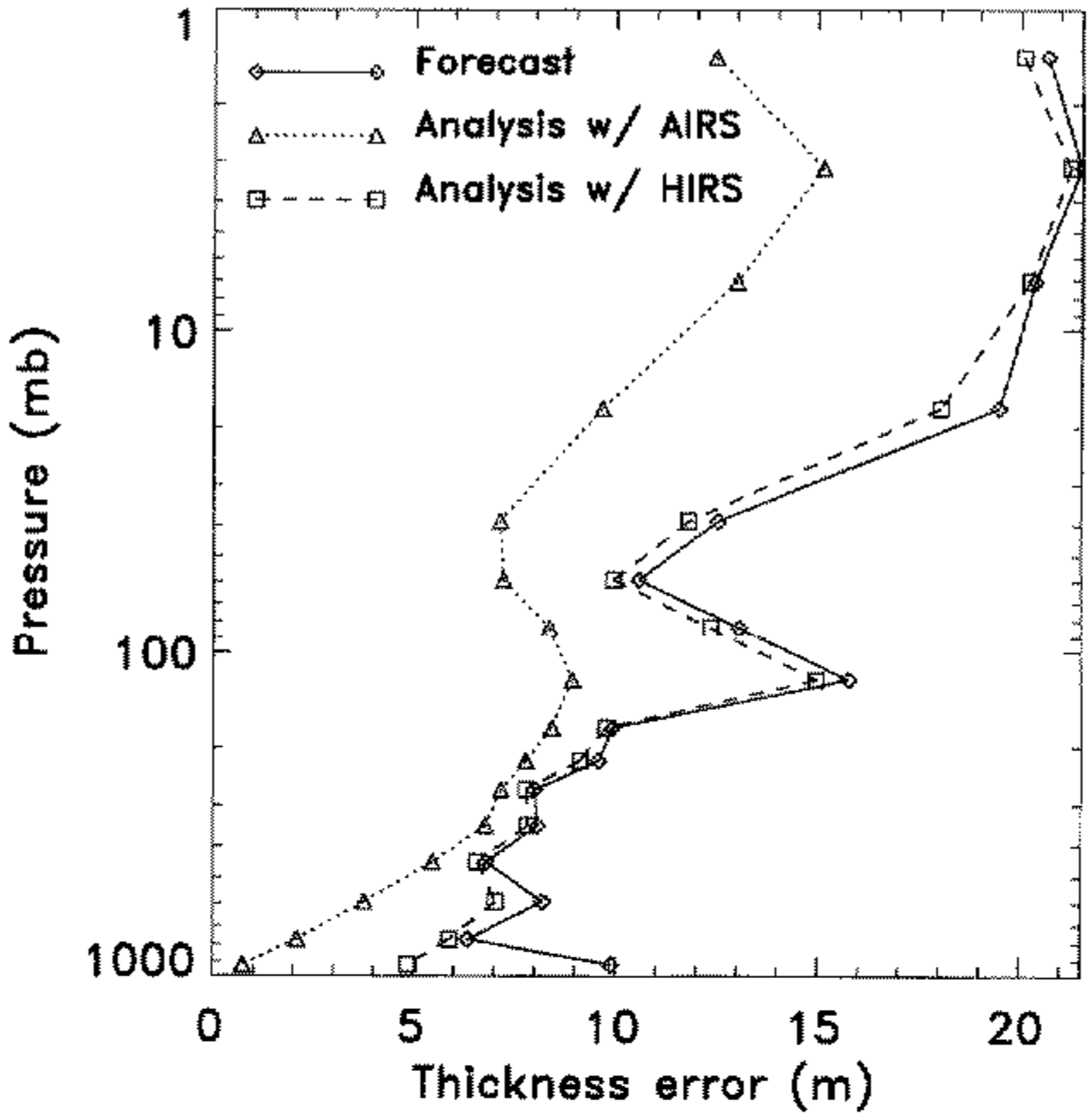

Figure 4. Expected thickness standard deviations (in layers bounded by the 18 Goddard Earth Observing System (GEOS) pressure levels plotted at the layer midpoint) computed using linear estimation theory for the midlatitude profile. See text for further explanation.

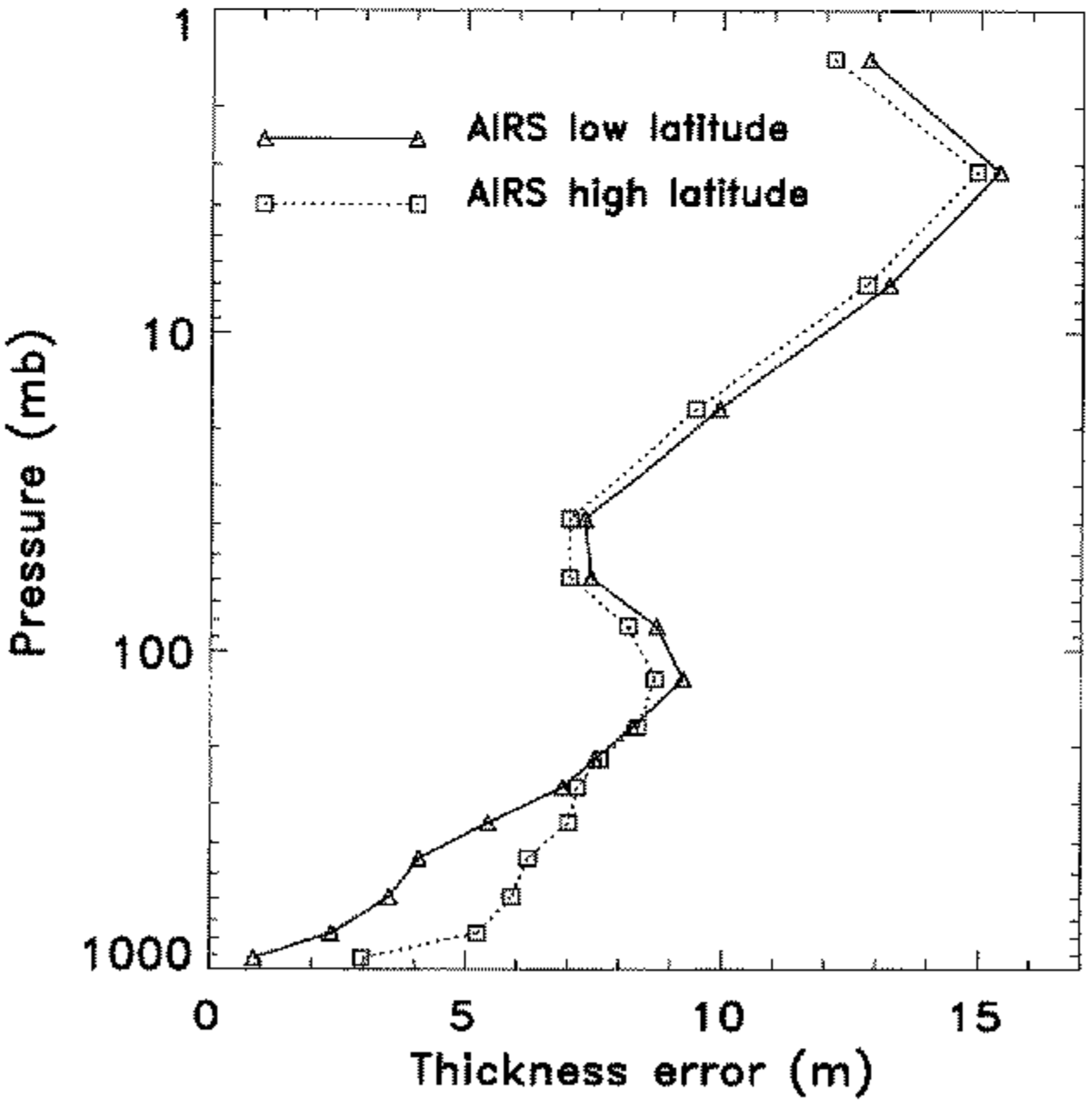

Figure 5. As Fig. 4 but showing the state dependence of the expected thickness errors.

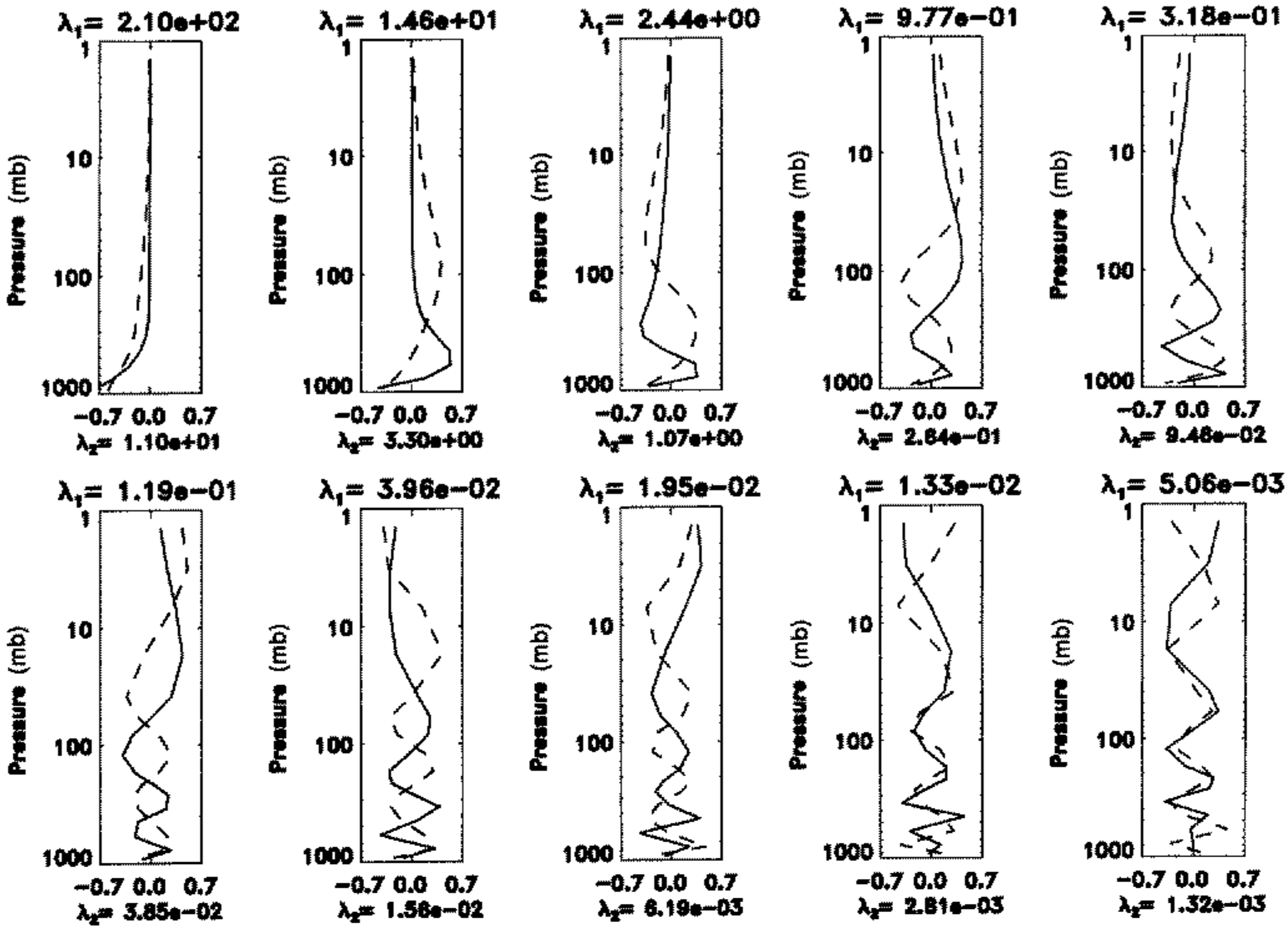

Figure 6. Leading eigenvectors and eigenvalues of $\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z$ for AIRS channels and the low-latitude profile (solid curve, λ_1) and high-latitude profile (dashed curve, λ_2). See text for explanation.

greater than that shown here. It should also be noted that the impact shown here is for a single sounding per grid-box. Analysing more soundings per grid box may further reduce analysis errors. On the other hand, analysis errors will increase (especially in the lowest level) when water vapour and surface parameters are estimated rather than taken to be the truth.

(*d*) *PED retrieval assimilation: Monte Carlo experiment*

Recall that PED retrieval assimilation involves assimilating the coefficients of the eigenvectors of $\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z$. Figures 6 and 7 show the leading eigenvectors and eigenvalues of $\mathbf{F}_z^{\mathrm{T}}(\mathbf{R}^y)^{-1}\mathbf{F}_z$ for AIRS and HIRS, respectively. The first few modes are very similar for the two instruments, but the error variances (inverse of the eigenvalues) are significantly lower for AIRS than for HIRS2. AIRS has a higher effective signal-to-noise ratio, primarily as a result of the larger number of channels.

We now examine the impact of a particular mode on the GEOS DAS in 1D. The quantity $1 - p_\ell^{\mathrm{a}}/p_\ell^{\mathrm{f}}$ is plotted as a function of mode ℓ in Fig. 8. AIRS radiances are shown to impact significantly on approximately 11 or 12 modes, while HIRS2 radiances impact on only about 4 or 5. While this type of analysis tells us how many modes the data have significant impact on, it does not provide any information about the magnitude of the analysis errors or relative importance of each mode. Figure 9 shows a measure of the total error reduction, trace($\mathbf{P}^{\mathrm{a}}$)/trace($\mathbf{P}^{\mathrm{f}}$), plotted as a function of the number of modes included in the analysis. Approximately 8–10 modes provide most of the AIRS impact, while approximately 4 modes account for the majority of the HIRS2 impact. Again, the total impact from AIRS is shown to be significantly greater than that from HIRS2.

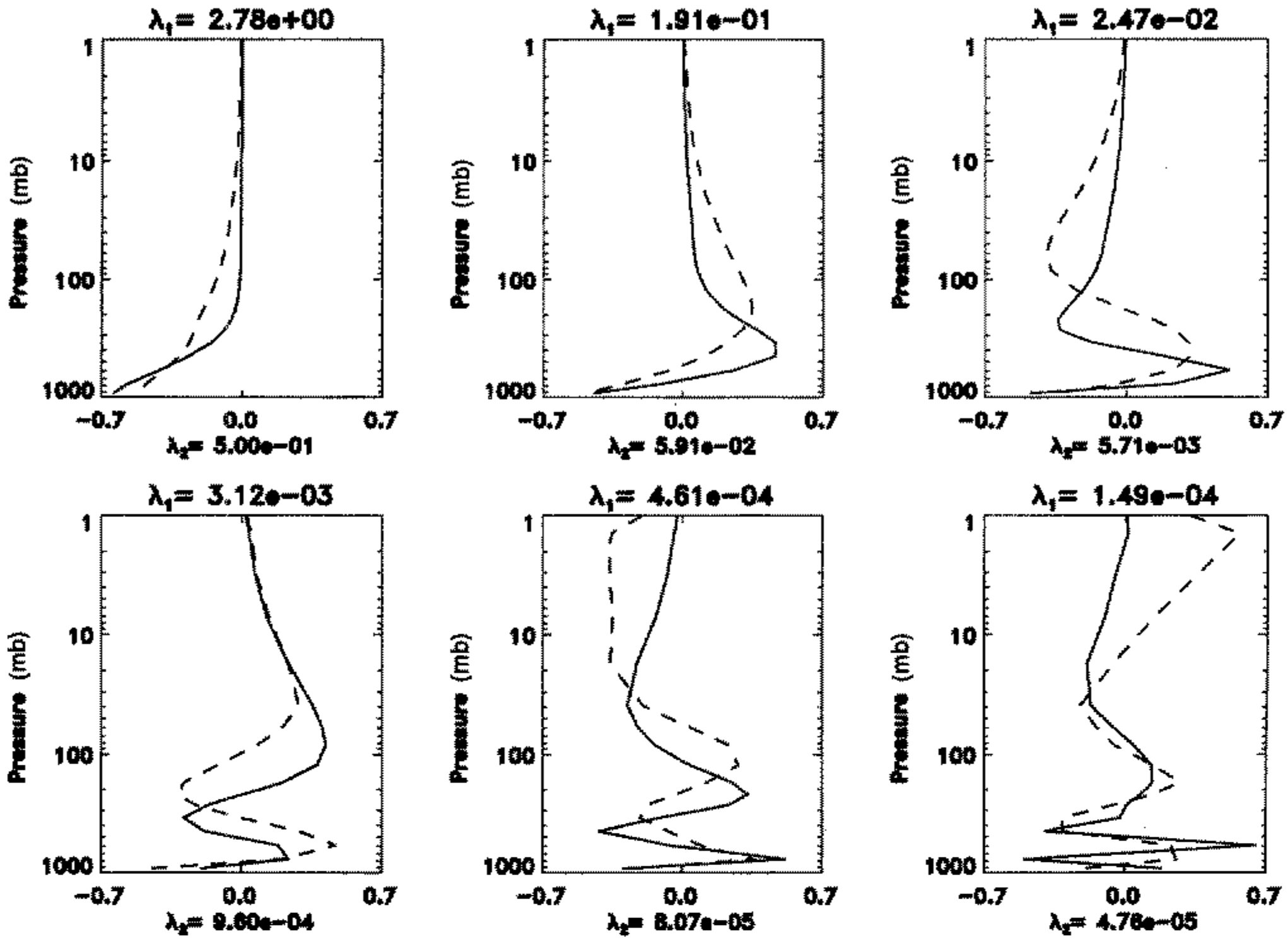

Figure 7. As Fig. 6 but for HIRS2 channels.

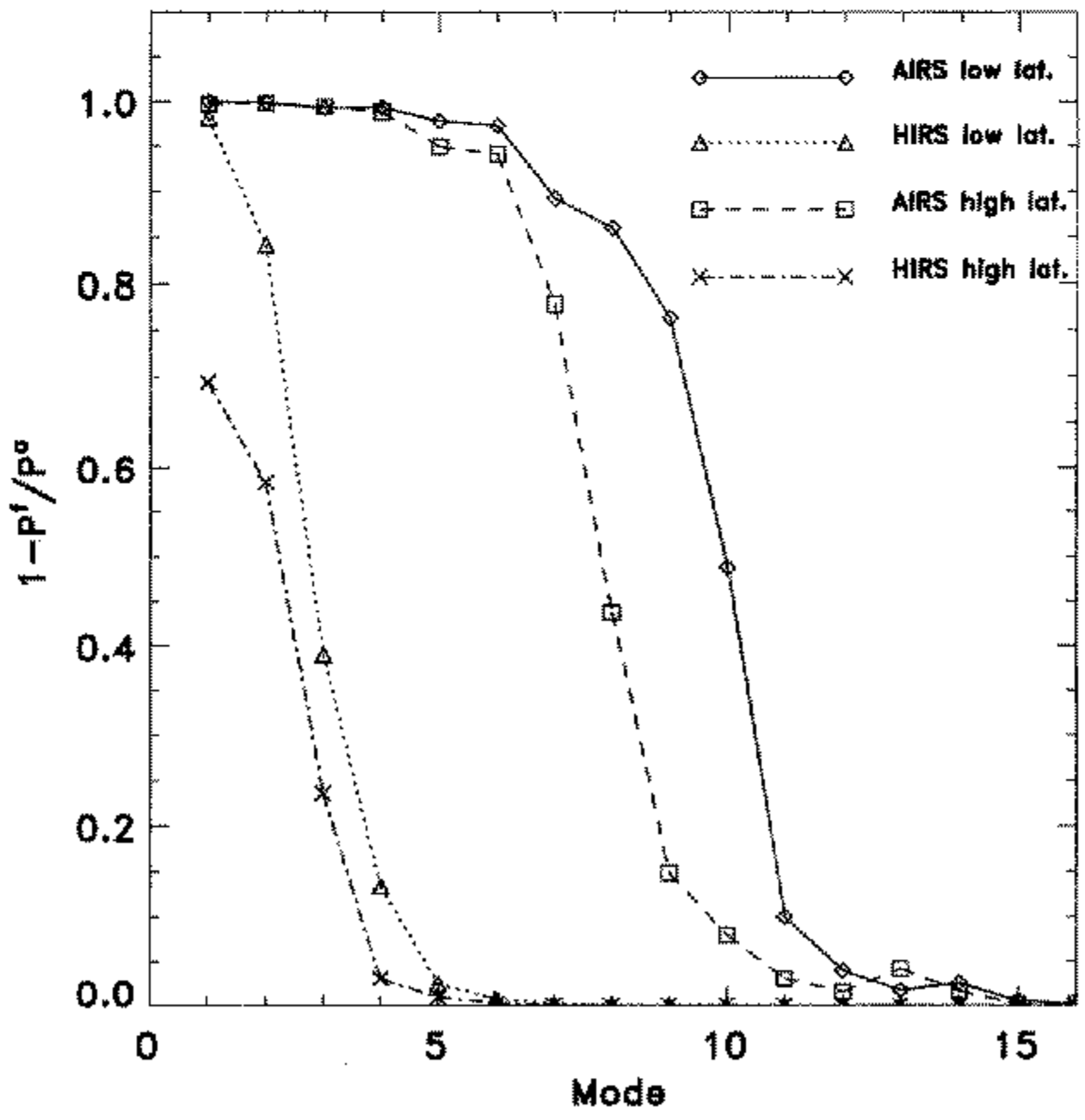

Figure 8. Fractional impact of data ($1\text{-}p_\ell^{\mathrm{f}}/p_\ell^{\mathrm{a}}$) for a given mode ℓ. See text for explanation.

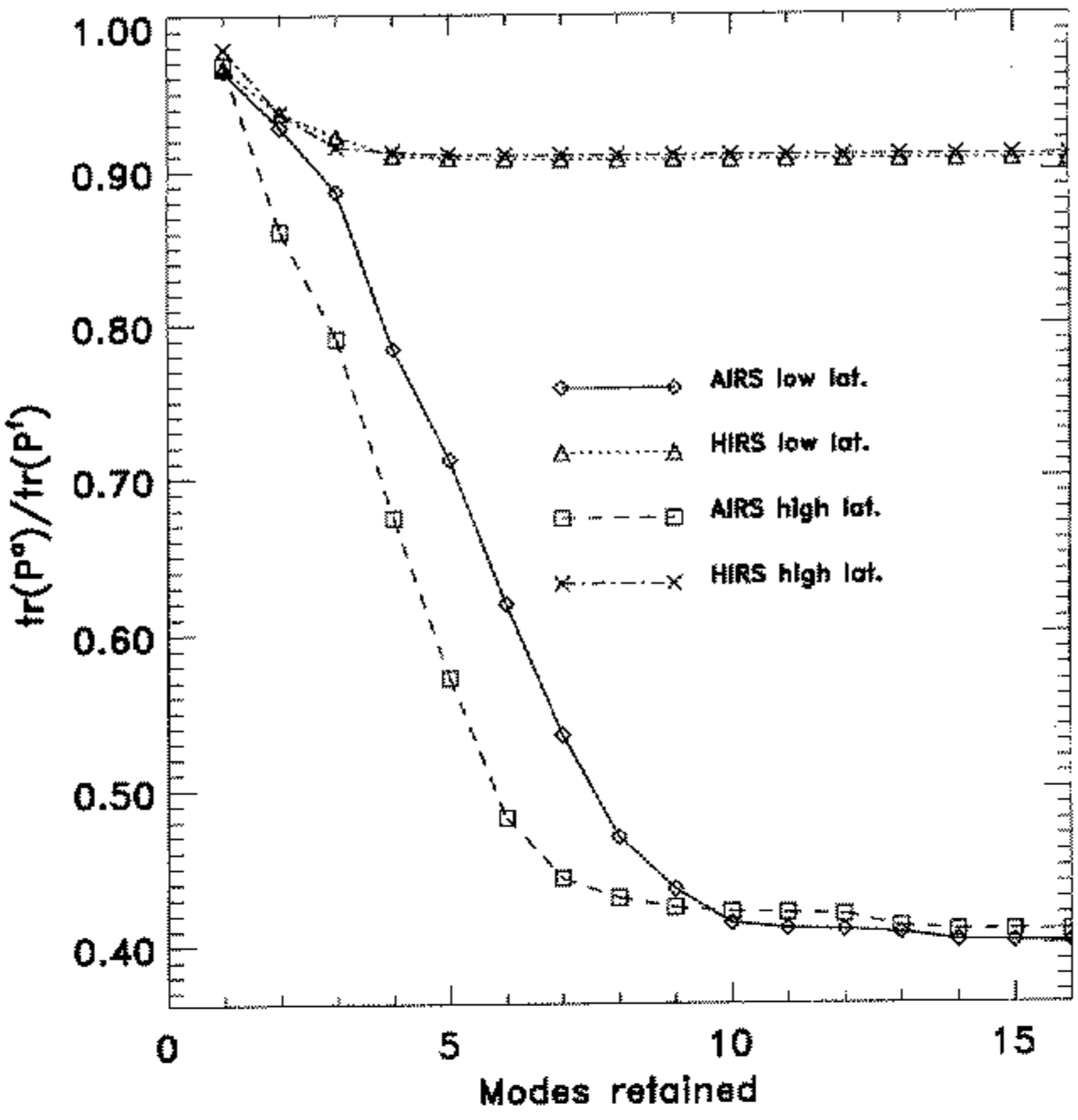

Figure 9. Measure of the reduction in analysis error, trace($\mathbf{P}^a$)/trace($\mathbf{P}^f$), as a function of number of modes included in the analysis. See text for further explanation.

Figure 10 shows the thickness errors of the Monte Carlo experiment for the AIRS high-latitude PED retrieval assimilation with 8, 9, and 10 modes included. The direct-radiance assimilation is also shown and is the optimal solution. We allowed three iterations in the 1D minimum variance retrieval. Further iterations did not significantly improve the results. The differences between the radiance assimilation and the PED retrieval assimilation are very small. Slight degradation is shown in the upper stratosphere when only 8–10 modes are assimilated, and the degradation increases slightly when the number of retained modes decreases.

Figure 11 is similar to Fig. 10, but shows results for the low-latitude case. Slightly more degradation is present when 8–10 modes are retained as compared with the high-latitude case. This result is expected, on the basis that Fig. 8 indicates more modes are needed for the low-latitude profile. Again, the difference between the 8–10 mode PED retrieval assimilation and radiance assimilation is very small. In the stratosphere, we have the unexpected result of the 8 mode PED assimilation outperforming the radiance assimilation. This has been attributed to linearization error that is not perfectly accounted for in $\mathbf{R}^y$. As verification, analyses were produced in Monte Carlo mode by linearizing the Jacobian about the truth, as shown in Fig. 12. The expected results are now obtained, and analysis errors are now nearly identical to those predicted by linear theory. This example shows that linearization error can be significant, and if not properly accounted for can degrade the analysis.

In this simulation, 550 AIRS radiance observations have been compressed into approximately 10 pieces of information. This would reduce the computation required by the PSAS conjugate gradient solver by approximately a factor of 55^2 or 3025 as compared with radiance assimilation. The data compression for HIRS2 is much smaller (approximately 10 channels to 4 pieces of information). It should be noted that the AIRS instrument, and instruments with similar spectral resolution such as the Infrared Atmospheric Sounding

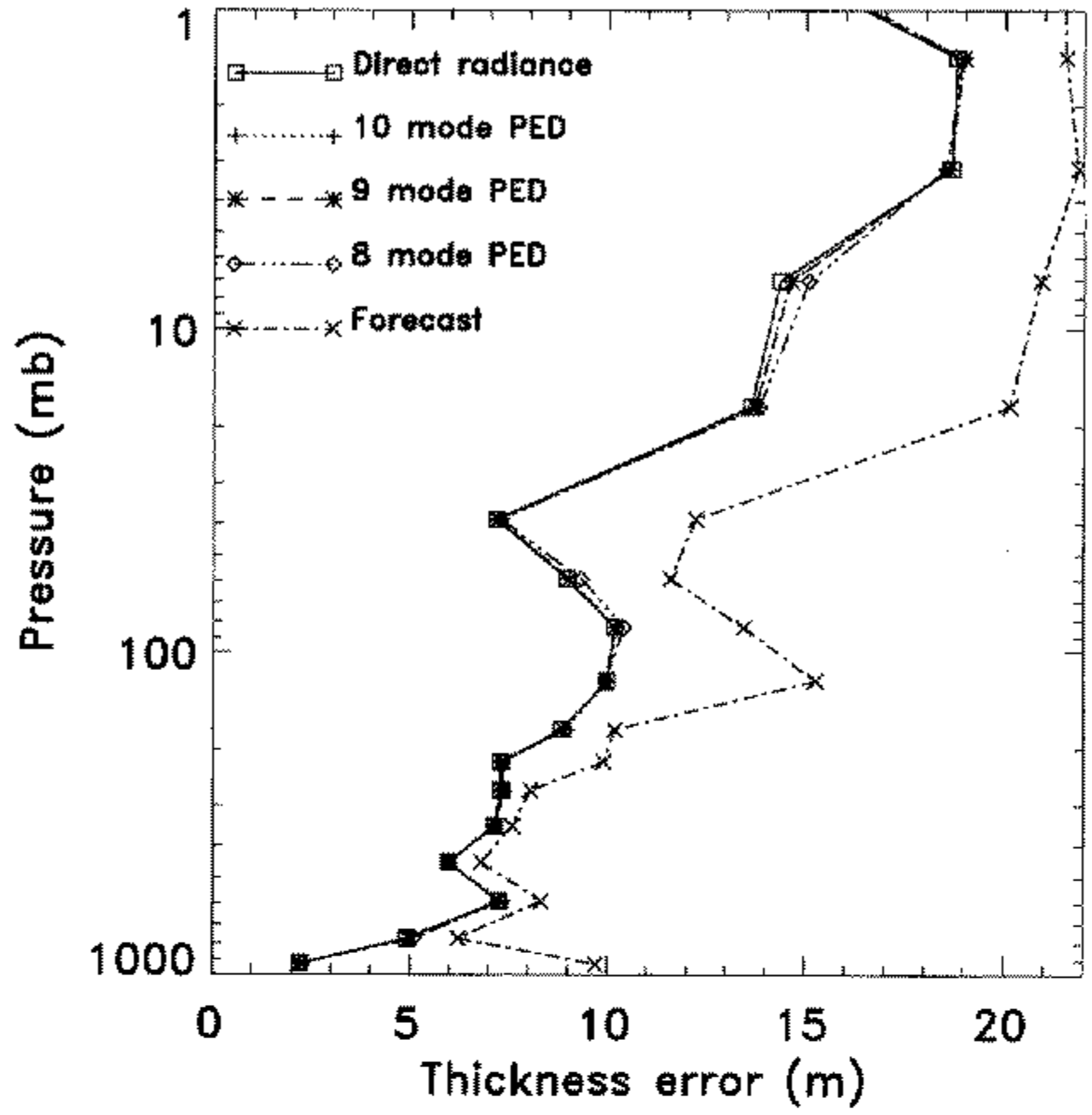

Figure 10. Thickness errors as in Fig. 4 obtained in the Monte Carlo experiment for radiance and partial eigendecomposition (PED) retrieval assimilation with Atmospheric Infrared Sounder (AIRS) channels for the high-latitude profile.

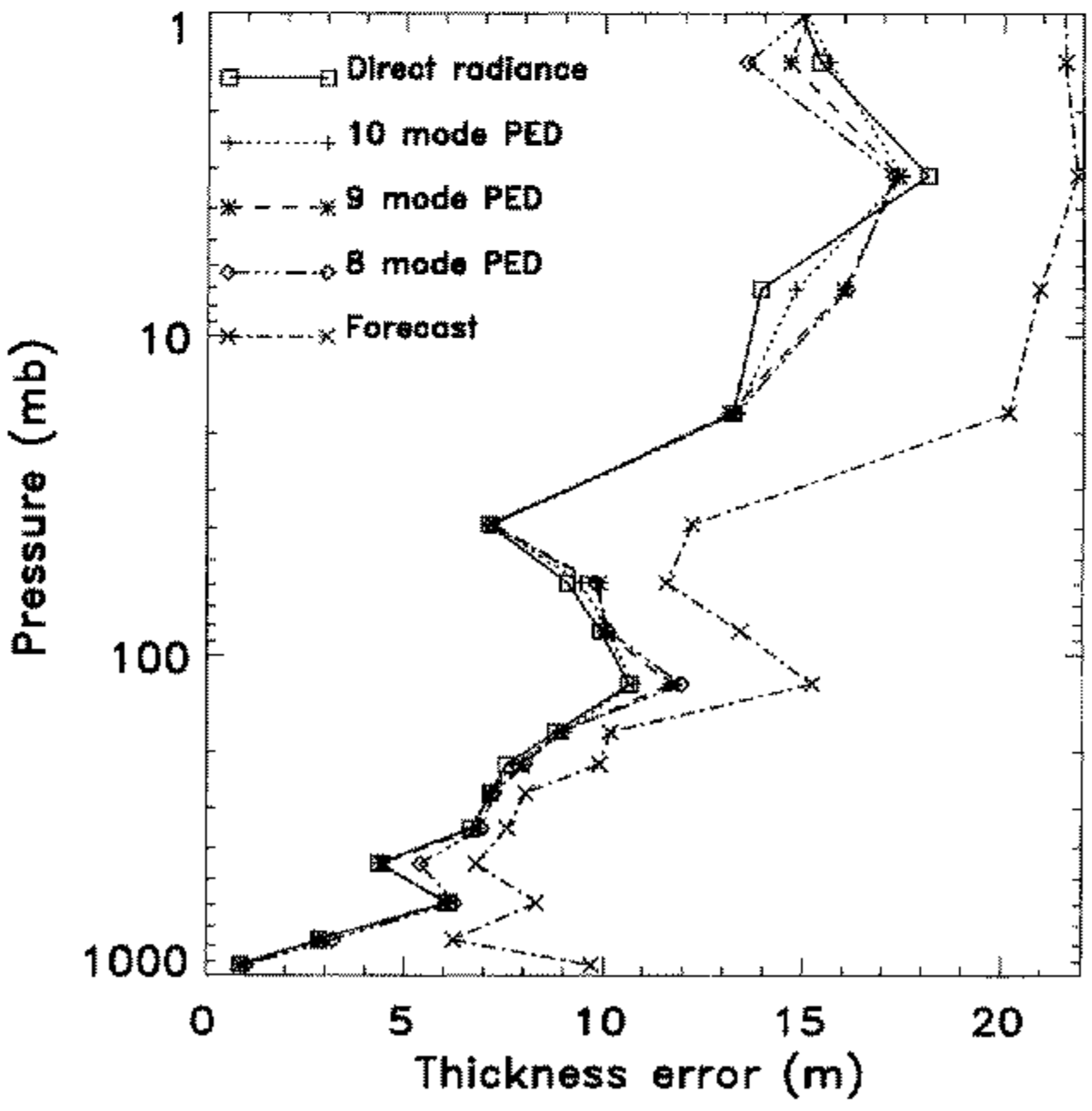

Figure 11. As Fig. 10 but for the low-latitude profile.

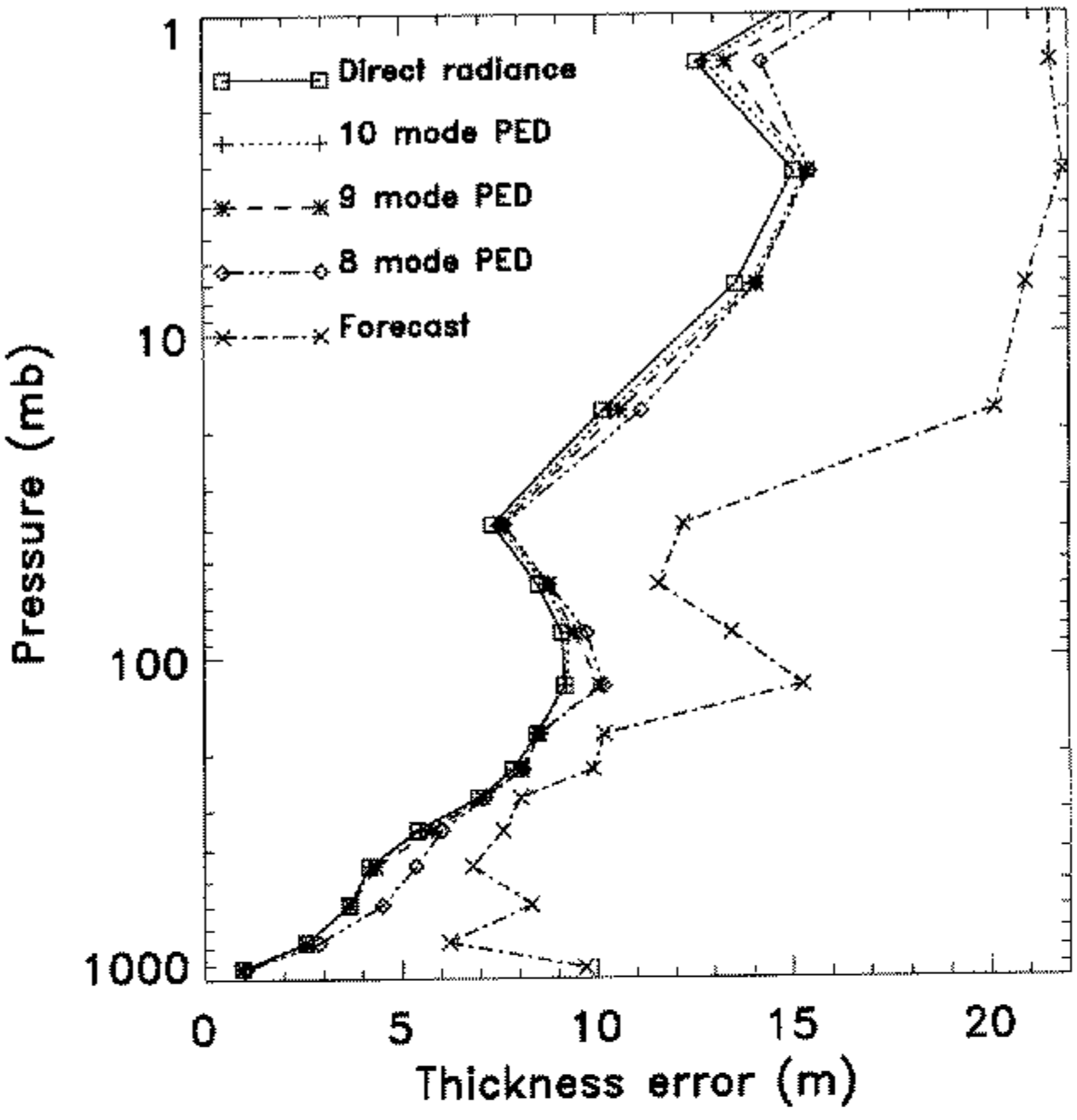

Figure 12. As Fig. 11 when linearization error is eliminated.

Interferometer (IASI), will provide many more than the 10 pieces of information shown here. We have considered here only temperature sounding and only a limited spectral region. AIRS and IASI will also provide information about the humidity profile, ozone profile, surface temperature, surface and cloud spectral emissivity, cloud height, and cloud fraction, that can potentially be used in a DAS. It has also been suggested that including channels in the 6.7μm water vapour band provides additional information about the temperature profile (Rodgers 1996). We note that to achieve this result, spectroscopic errors will have to be reduced from their current levels.

(e) *NSF retrieval assimilation*

Figure 13 shows the trailing modes of $(\mathbf{A}-\mathbf{I})$ and $\mathbf{R}^s=(\mathbf{A}-\mathbf{I})\mathbf{P}^p(\mathbf{A}-\mathbf{I})^T$ that are used in NSF methods 1 and 2, respectively, for AIRS and the low-latitude profile. In this example, $\mathbf{P}^p$ is taken to be $\mathbf{P}^f$ as would be the case for interactive retrievals. $\mathbf{D}_y$ is assumed to be that from a minimum-variance retrieval. These modes, as expected, are similar to each other and similar to the leading modes of $\mathbf{F}_z(\mathbf{R}^y)^{-1}\mathbf{F}_z$ shown in Fig. 6. Figure 14 is similar to Fig. 10, but shows the results for the two methods of NSF retrieval assimilation using different numbers of modes. The difference between these approaches and radiance assimilation is insignificant, especially when 10 modes are retained. The results of the NSF and PED analyses are nearly identical even though the PED approach is more optimal from a theoretical point of view. However, it remains to be seen how well the NSF approach would perform when less optimal prior information is used in the retrieval. As with PED retrieval assimilation, the computation required for full 3D PSAS will be reduced using NSF retrieval assimilation as compared with radiance assimilation, especially for AIRS and IASI.

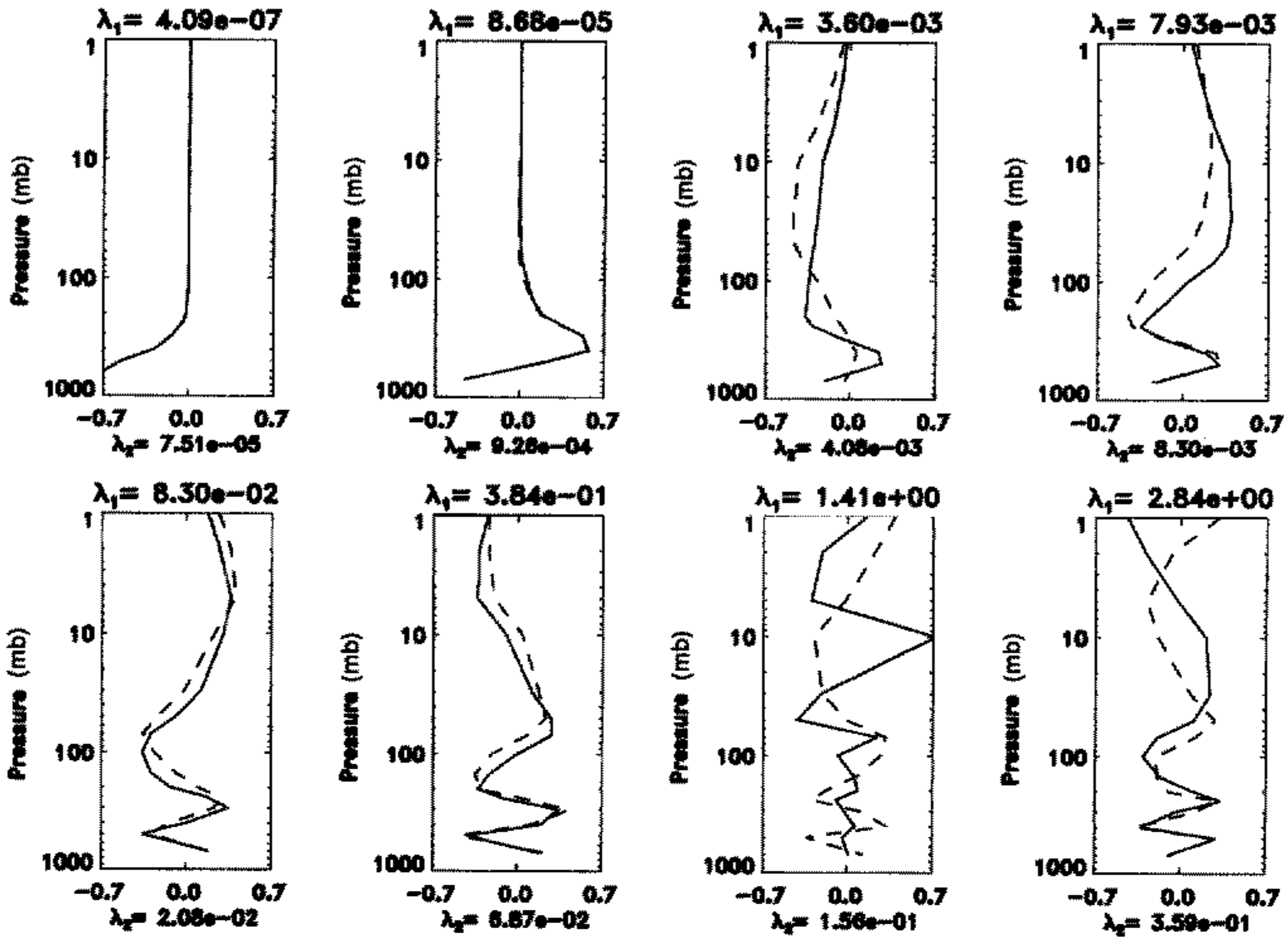

Figure 13. Trailing modes of $\mathbf{A}-\mathbf{I}$ (dashed line) with singular values λ_2 and trailing modes of $\mathbf{R}^s$ (solid line) with eigenvalues λ_1 using AIRS channels and low-latitude profile. See text for further explanation.

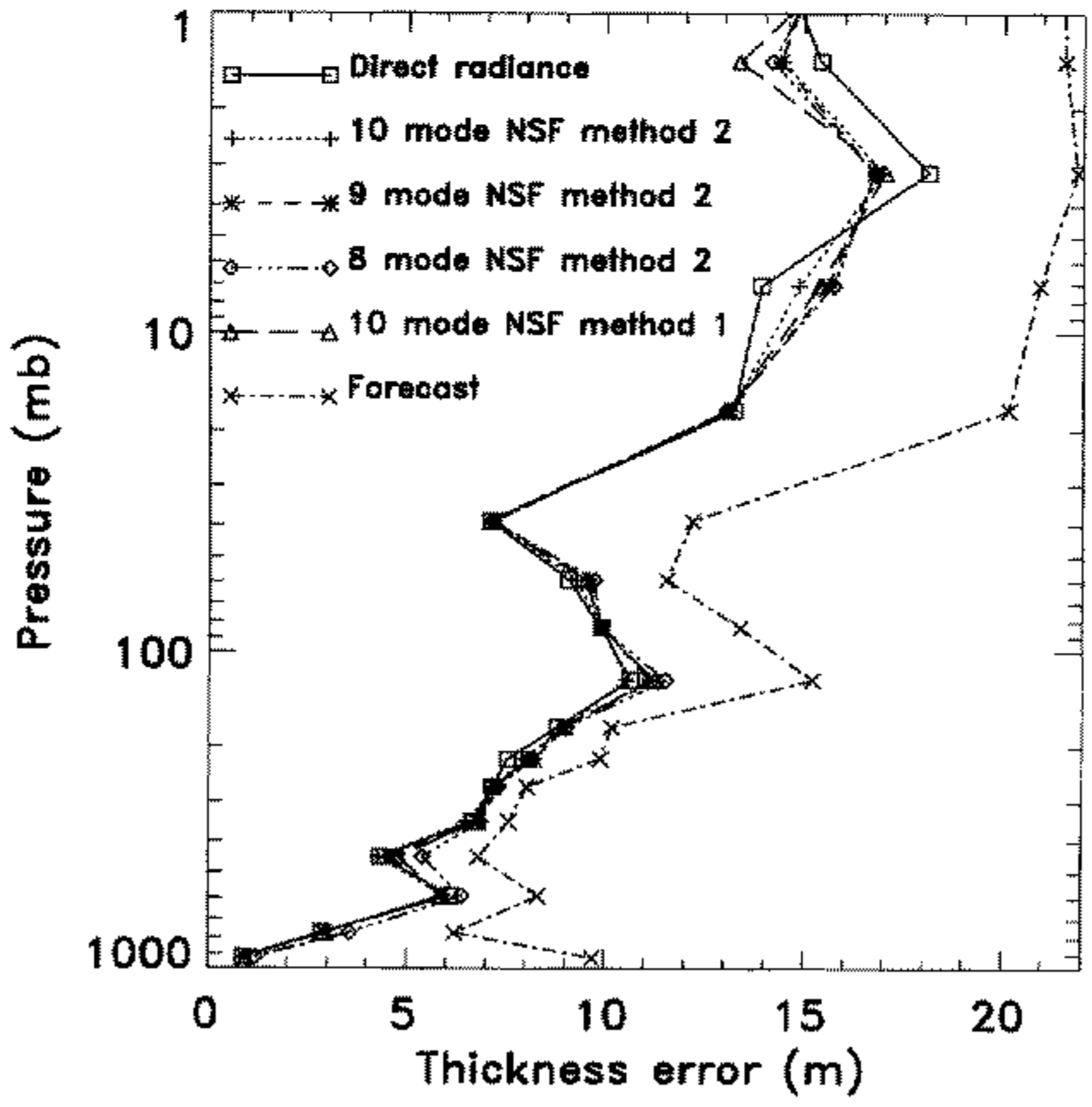

Figure 14. Similar to Fig. 11 but showing errors for null-space filtered (NSF) retrieval assimilation.

8. Conclusions and future work

We have presented the theoretical basis for consistently assimilating remotely sensed data in the form of PED (partial eigen-decomposition) and NSF (null-space filtered) retrievals. In this approach effects of the prior information are reduced or eliminated from the retrievals, so that the assumption of zero correlation between the retrieval and forecast errors is justified. In addition, the assimilation methods developed here compress data, and can reduce the amount of computation needed in some DAS schemes while preserving essentially all of the information content of radiance measurements. In a 1D PSAS simulation these techniques produced results nearly identical to radiance assimilation. In a 3D system there is the potential for a significant reduction in computation, especially for future high-spectral-resolution instruments. Performing the expensive radiative-transfer calculations off-line once and for all using the approaches developed here would also significantly reduce costs when performing parallel system tests, such as observation system experiments and observation system simulation experiments, or when running several multiple-year re-analyses.

Future plans include implementation of either (or both) the PED or NSF approaches in a PSAS using data from TOVS as a prototype for AIRS and IASI. Several implementation issues remain to be addressed, including whether it is possible to select a fixed or small subset of transformations in a more nonlinear situation (i.e. when humidity and other constituents are included). The success of these methods will be evaluated both in terms of the cost and quality of the analyses. The success of any chosen method will of course depend upon the ability to remove systematic errors, accurately estimate observation and background errors, and apply effective quality control to the observations. We also plan to investigate the application of these methods to the assimilation of retrieved ocean surface winds, cloud top heights, and precipitation.

Acknowledgements

The authors are grateful to R. Ménard, S. E. Cohn, D. P. Dee, C. D. Rodgers, R. J. Purser, L. Rokke, and two anonymous reviewers for enlightening discussions and helpful comments that contributed to this paper.

References

Andersson, E., Pailleux, J., Thépaut, J. N., Eyre, J. R., McNally A. P., Kelly, G. A., and Courtier, P. 1994 Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **120**, 627–653

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G. A., Vasiljević, D., Branković, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J.-N., and Viterbo, P. 1998 The ECMWF implementation of three dimensional variational assimilation (3D-Var). III: Experimental results. *Q. J. R. Meteorol. Soc.*, **124**, in press

Backus, G. E. and Gilbert, J. F. 1970 Uniqueness in the inversion of inaccurate gross earth data, *Philos. Trans. R. Soc. London*, **A266**, 123–192

Bengtsson, L. and Shukla, J. 1988 Integration of space and in situ observations to study global climate change. *Bull. Am. Meteorol. Soc.*, **69,** 1130–1143

Chédin A., Scott, N. A., Wahiche, C. and Moulinier, P. 1985 The improved initialisation inversion method: a high resolution physical method for temperature retrievals from the TIROS-N series. *J. Clim. Appl. Meteorol.*, **24,** 128–143

Courtier, P., Andersson, E., Heckley, W., Kelly, G., Pailleux, J., Rabier, F., Thepaut, J.-N., Unden, P., Vasiljevic, D., Cardinali, C., Eyre, J. R., Hamrud, M., Haseler, J., Hollingsworth, A., McNally, A. P. and Stoffelen, A. 1993 'Variational Assimilation at ECMWF'. ECMWF Tech. Memo. 194, Reading, UK

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljević, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M. 1998 The ECMWF implementation of three dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.*, **124,** in press

Daley, R. 1991 *Atmospheric data analysis*. Cambridge University Press, New York

1992 The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Weather Rev.*, **120,** 197–207

Dee, D. P. 1995 On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.*, **123,** 1128–1145.

Derber, J. and Wu, W.-S. 1996 'The use of cloud-cleared radiances in the NCEP's SSI analysis system'. Pp 236–237 in Proceedings of the 11th Conference of the American Meteorological Society on Numerical Weather Prediction, Norfolk, USA

Escobar-Munoz, J., Chédin, A., Cheruy, F. and Scott, N. A. 1993 Réseaux de neurones multi-couches pour la restitution de variables thermodynamiques atmosphériques à l'aide de sonduers verticaux satellitaires. *C.R.A.S.*, **317**, 911–918

Eyre, J. R. 1987 On systematic errors in satellite sounding products and their climatological mean values. *Q. J. R. Meteorol. Soc.*, **113,** 279–292

1990 The information content of data from satellite sounding systems: A simulation study. *Q. J. R. Meteorol. Soc.*, **116,** 401–434

1992 'A bias correction scheme for simulated TOVS brightness temperatures'. ECMWF Tech. Memo. 186, Reading, UK

Eyre, J. R., Kelly, G. A., McNally, A. P., Andersson, E. and Persson, A. 1993 Assimilation of TOVS radiance information through one-dimensional variational analysis. *Q. J. R. Meteorol. Soc.*, **119,** 1427–1463

Gibson, J. K., Hernandez, A., Kallberg, P., Nomura, A. Serrano, E. and Uppala, S. 1994 'The ECMWF re-analysis project'. Pp 288–291 in the Preprints of the Tenth Conference of the American Meteorological Society on Numerical Weather Prediction, Portland, USA

Goldberg, M. D., Fleming, H. E., Baker, W. E. and Derber, J. C. 1993 'The development of an interactive retrieval–analysis–forecast algorithm'. Pp 163–168 in Technical Proceedings of the 7th International TOVS study conference. Ed. J. Eyre. Igls, Austria

Guo, J. and da Silva, A. M. 1995 'Computational aspects of Goddard's physical-space statistical analysis system (PSAS)'. In Preprints of Second UNAM-CRAY Supercomputing Conference on Numerical Simulations in the Environmental and Earth Sciences. Mexico City, Mexico

Hotelling, H. 1933 Analysis of complex statistical variables into principle components. *J. Educ. Psychol.* **24,** 417–441 and 498–520

Huang, H. L. and Purser, R. J. 1996 Objective measures of the information density of satellite data. *Meteorol. Atmos. Phys.*, **60,** 105–117

Jazwinski, A. H. 1970 *Stochastic processes and filtering theory.* Academic Press, New York, USA

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Leetma, A., Reynolds, R., Jenne, R. and Joseph, D. 1995 The NMC/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.*, **77,** 437–471

Karhunen, K. 1947 Uber lineare methoden in der wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae*, **A137**

Koschman, A. 1954 'On the filtering of nonstationary time series'. P. 126 in Proceedings of the 1954 National Electronic Conference

Loeve, M. 1948 'Fonctions Aleatoires de second order'. In *Processus Stochastiques et Mouvement Brownien*, Ed. P. Levy. Hermann, Paris, France

Lorenc, A. C. 1981 A global three-dimensional multivariate statistical interpolation scheme. *Mon. Weather Rev.*, **109,** 701–721

1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112,** 1177–1194

Lorenc, A. C., Adams, W. and Eyre, J. R. 1986 'The analysis of high-resolution satellite data by the Meteorological Office'. Pp. 69–87 in the Proceedings of the ECMWF Workshop on high-resolution analysis. 24–26 June 1985, Reading, UK

Navon, I. M. and Legler, D. 1987 Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Weather Rev.*, **115,** 1479–1502

Parrish, D. F. and Derber, J. C. 1992 The National Meteorological Center's statistical spectral interpolation analysis system. *Mon. Weather Rev.*, **120,** 1747–1763

Prunet, P., Thépaut, J.-N. and Cassé, V. 1998 The information content of clear-sky IASI radiances and their potential for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **124,** 211–241

Purser, R. J. 1990 'Vertical aspects of the assimilation of sounding data'. Pp. 501–505 in Preprints of the WMO International Symposium on assimilation of observations in meteorology/remote sensing and applications, 13–16 May 1986. American Meteorological Society, Williamsburg, USA

Purser R. J. and Huang, H. L. 1993 Estimating the effective data density in a satellite retrieval or an objective analysis. *J. Appl. Meteorol.*, **32,** 1092–1107

Rabier, F., McNally, A. P., Courtier, P., Undén, P., Eyre, J. R., Hollingsworth, A. and Bouttier, F. 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Q. J. R. Meteorol. Soc.*, **124,** in press

Rodgers, C. D. 1976 Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. Space Phys.*, **14,** 609–624

1990 Characterization and error analysis of profiles retrieved from remote sounding measurements. *J. Geophys. Res.*, **95,** 5587–5595

1996 'Information content and optimisation of high spectral resolution measurements'. Pp. 136–147 in *Optical spectroscopic techniques and instrumentation for atmospheric and space research II.* Eds. Paul B. Hays and Jinxue Wang. SPIE, Denver, USA

Schubert, S. D., Rood, R. B. and Pfaendtner, J. 1993 An assimilated dataset for earth sciences applications. *Bull. Am. Meteorol. Soc.*, **74,** 2331–2342

da Silva, A. M., Pfaendtner, J., Guo, J., Sienkiewicz, M. and Cohn, S. E. 1995 'Assessing the effects of data selection with DAO's Physical-space Statistical Analysis System'. Pp. 273–278 in Proceedings of the International Symposium on assimilation of observations in meteorology and oceanography. Tokyo, Japan

da Silva, A. M., Redder, C. and Dee, D. P. 1996 'Modeling retrieval error covariances for global data assimilation'. Pp. 503–507 in Proceedings of the 8th Conference on satellite meteorology. Atlanta, USA

Smith, W. L. and Woolf, H. 1976 The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations. *J. Atmos. Sci.*, **33,** 1–7

Smith, W. L., Woolf, H. M., Hayden, C. M., Wark, D. Q. and McMillin, L. M. 1979 The TIROS-N operational vertical sounder. *Bull. Am. Meteorol. Soc.*, **60,** 1177–1187

Sullivan, J., Gandin, L., Gruber, A. and Baker, W. 1993 Observation error statistics for NOAA-10 temperature and height retrievals. *Mon. Weather Rev.*, **121,** 2578–2587

Susskind, J. and Pfaendtner, J. 1989 'Impact of iterative physical retrievals on NWP'. Pp. 245–270 in Report on the Joint ECMWF/EUMETSAT Workshop on the use of satellite data in operational weather prediction: 1989–1993, Vol. 1. Ed. A. Hollingsworth, Reading, UK

Susskind, J., Rosenfield, J. and Reuter, D. 1983 An accurate radiative transfer model for use in the direct physical inversion of HIRS and MSU temperature sounding data. *J. Geophys. Res.* **88,** 8550–8568

Talagrand, O. 1988 'Four-dimentional variational assimilation'. Pp. 1–30 in Proceedings of the ECMWF Seminar on data assimilation and the use of satellite data, Vol. 2. Reading, UK

Thompson, O. 1992 Regularizing the satellite temperature-retrieval problem through singular-value decomposition of the radiative transfer physics. *Mon. Weather Rev.*, **120,** 2314–2328

Twomey, S. 1974 Information content in remote sensing. *Appl. Opt.*, **13**, 942–945